Application-driven Advances in Multi-biometric Fusion

dem Fachbereich Informatik
der Technischen Universität Darmstadt
vorzulegende

DISSERTATION

zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs (Dr.-Ing.)
von

M.Sc. Naser Damer

geboren in Amman, Jordanien

Referenten der Arbeit:	Prof. Dr. Arjan Kuijper Technische Universität Darmstadt
	Prof. Dr. Dieter W. Fellner Technische Universität Darmstadt
	Prof. Dr. Raghavendra Ramachandra Norwegian University of Science and Technology
Tag der Einreichung:	22/01/2018
Tag der mündlichen Prüfung:	05/03/2018

Darmstädter Dissertation
D 17

Damer, Naser: Application-driven Advances in Multi-biometric Fusion
Darmstadt, Technische Universität Darmstadt
Jahr der Veröffentlichung der Dissertation auf TUprints: 2018
URN: urn:nbn:de:tuda-tuprints-73248
Tag der mündlichen Prüfung: 05.03.2018

Erklärung zur Dissertation

Hiermit versichere ich die vorliegende Dissertation selbständig nur mit den angegebenen Quellen und Hilfsmitteln angefertigt zu haben. Alle Stellen, die aus Quellen entnommen wurden, sind als solche kenntlich gemacht. Diese Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

Darmstadt, den 22/01/2018 Naser Damer

Abstract

Biometric recognition is the automated recognition of individuals based on their behavioral or biological characteristics. Beside forensic applications, this technology aims at replacing the outdated and attack prone, physical and knowledge-based, proofs of identity. Choosing one biometric characteristic is a tradeoff between universality, acceptability, and permanence, among other factors. Moreover, the accuracy cap of the chosen characteristic may limit the scalability and usability for some applications. The use of multiple biometric sources within a unified frame, i.e. multi-biometrics, aspires to tackle the limitations of single source biometrics and thus enables a wider implementation of the technology. This work aims at presenting application-driven advances in multi-biometrics by addressing different elements of the multi-biometric system work-flow.

At first, practical oriented pre-fusion issues regarding missing data imputation and score normalization are discussed. This includes presenting a novel performance anchored score normalization technique that aligns certain performance-related score values in the fused biometric sources leading to more accurate multi-biometric decisions when compared to conventional normalization approaches. Missing data imputation within score-level multi-biometric fusion is also addressed by analyzing the behavior of different approaches under different operational scenarios.

Within the multi-biometric fusion process, different information sources can have different degrees of reliability. This is usually influenced in the fusion process by assigning relative weights to the fused sources. This work presents a number of weighting approaches aiming at optimizing the decision made by the multi-biometric system. First, weights that try to capture the overall performance of the biometric source, as well as an indication of its confidence, are proposed and proved to outperform the state-of-the-art weighting approaches. The work also introduces a set of weights derived from the identification performance representation, the cumulative match characteristics. The effect of these weights is analyzed under the verification and identification scenarios.

To further optimize the multi-biometric process, information besides the similarity between two biometric captures can be considered. Previously, the quality measures of biometric captures were successfully integrated, which requires accessing and processing raw captures. In this work, supplementary information that can be reasoned from the comparison scores are in focus. First, the relative relation between different biometric comparisons is discussed and integrated in the fusion process resulting in a large reduction in the error rates. Secondly, the coherence between scores of multi-biometric sources in the same comparison is defined and integrated into the fusion process leading to a reduction in the error rates, especially when processing noisy data.

Large-scale biometric deployments are faced by the huge computational costs of running biometric searches and duplicate enrollment checks. Data indexing can limit the search domain leading to faster searches. Multi-biometrics provides richer information that can enhance the retrieval performance. This work provides an optimizable and configurable multi-biometric data retrieval solution that combines and enhances the robustness of rank-level solutions and the performance of feature-level solutions.

Furthermore, this work presents biometric solutions that complement and utilize multi-biometric fusion. The first solution captures behavioral and physical biometric characteristics to assure a continuous user authentication. Later, the practical use of presentation attack detection is discussed by investigating the more realistic scenario of cross-database evaluation and presenting a state-of-the-art performance comparison. Finally, the use of multi-biometric fusion to create face references from videos is addressed. Face selection, feature-level fusion, and score-level fusion approaches are evaluated under the scenario of face recognition in videos.

Zusammenfassung

Biometrische Erkennung bezeichnet die automatische Erkennung von Personen auf Grundlage ihres Verhaltens oder ihrer biologischen Eigenschaften. Neben forensischen Anwendungen zielt diese Technologie darauf ab, die traditionellen, wissensbasierten und angriffsanfälligen Identitätsnachweise zu ersetzen. Die Wahl eines biometrischen Merkmals ist neben anderen Faktoren vor allem ein Kompromiss zwischen Universalität, Akzeptanz und Dauerhaftigkeit. Darüber hinaus kann die Genauigkeitsobergrenze des gewählten Merkmals die Skalierbarkeit und Nutzbarkeit für einige Anwendungen einschränken. Die Nutzung mehrerer biometrischer Merkmale innerhalb eines einheitlichen Rahmens (Multi-Biometrie) zielt darauf ab, die Beschränkungen der Biometrie einzelner Merkmale anzugehen und somit eine breitere Implementierung der Technologie zu ermöglichen. Diese Arbeit präsentiert Fortschritte in der Multi-Biometrie, indem sie verschiedene Schritte des multi-biometrischen Workflows genauer untersucht.

Zunächst werden praxisorientierte Fragen in Bezug auf Imputation fehlender Daten und Vergleichswert-Normalisierung vor der Fusion diskutiert. Dies umfasst die Vorstellung einer neuartigen leistungsverankerten Vergleichswert-Normalisierungstechnik, die bestimmte leistungsbezogene Vergleichswerte in den zu fusionierenden biometrischen Quellen ausrichtet, was zu einer präziseren multi-biometrischen Entscheidungsfindung im Vergleich zu herkömmlichen Normalisierungsansätzen führt. Ebenfalls wird die Imputation fehlender Daten innerhalb der wertebasierenden multi-biometrischen Fusion durch die Analyse des Verhaltens verschiedener Ansätze in verschiedenen operativen Szenarien untersucht.

Innerhalb des multi-biometrischen Fusionsprozesses können verschiedene Informationsquellen unterschiedliche Zuverlässigkeitsgrade aufweisen. Dies wird in der Regel durch die Zuordnung von relativen Gewichten zu den zu fusionierenden Quellen in den Fusionsprozess erreicht. Diese Arbeit präsentiert eine Reihe von Gewichtungsansätzen zur Optimierung der Entscheidungsfindung des biometrischen Systems. Dabei werden Gewichte, die auch als Vertrauensindikator fungieren, zur Erfassung der Gesamtleistung der biometrischen Quelle vorgeschlagen und deren Überlegenheit gegenüber aktueller Gewichtungsansätze unter Beweis gestellt. Diese Arbeit führt ebenfalls einige Gewichte ein, die aus der Darstellung des Identifikationsleistung abgeleitet wurden. Die Auswirkungen dieser Gewichte werden sowohl im Verifikations- als auch im Identifikationsszenario analysiert.

Zur weiteren Optimierung des multi-biometrischen Prozesses können auch Informationen außer der Ähnlichkeit zweier biometrischer Aufnahmen berücksichtigt werden. Bisher wurden die Qualitätsmaße der biometrischen Aufnahmen erfolgreich integriert, wenn Rohdaten der Aufnahmen zur Verarbeitung zur Verfügung standen. In dieser Arbeit stehen ergänzende Informationen, die sich aus den Vergleichswerten ableiten lassen, im Fokus. Zunächst wird die relative Beziehung zwischen den verschiedenen biometrischen Vergleichen diskutiert und in den Fusionsprozess integriert, was zu einer erheblichen Reduzierung der Fehlerraten führt. Weiterhin wird die Kohärenz zwischen den Werten der multi-biometrischen Quellen im selben Vergleich definiert und in den Fusionsprozess integriert. Dies führt dazu, dass es zu einer weiteren Verringerung der Fehlerraten kommt, insbesondere bei der Verarbeitung verrauschter Daten.

Beim Einsatz großskalierter biometrischer Anwendungen werden diese mit enormen Kosten konfrontiert, die bei biometrischen Suchvorgängen und bei der doppelten Registrierungsprüfung entstehen. Die Datenindexierung kann die Suchdomäne einschränken, was zu schnelleren Suchvorgängen und verringerten Kosten führt. Dabei liefert die Multi-Biometrie umfangreichere Informationen, die die Suchleistung verbessern können. Diese Arbeit

bietet eine optimierbare und konfigurierbare multi-biometrische Datenabruflösung, die die Robustheit Rang-basierter und die Leistungsfähigkeit Feature-basierter Lösungen kombiniert und verbessert.

Darüber hinaus werden in dieser Arbeit neue biometrische Lösungen vorgestellt, die die multi-biometrische Fusion ergänzen und nutzen. Die erste Lösung erfasst verhaltensbezogene und physikalische biometrische Merkmale, um kontinuierliche Authentisierung zu gewährleisten. Später wird die praktische Anwendung der Erkennung von Präsentationsangriffen diskutiert, indem das realistischere Szenario der datenbankübergreifenden Auswertung untersucht wird. Weiterhin werden Leistungsvergleiche auf dem neuesten Stand der Technik präsentiert. Zum Abschluss wird die Verwendung der multi-biometrischen Fusion zur Erzeugung von Gesichtstemplates aus Videos diskutiert. Gesichtsauswahl, Fusion auf Feature-Ebene und Fusionsverfahren auf Score-Ebene werden im Rahmen des Szenarios der Gesichtserkennung in Videos ausgewertet, um die Vorteile der multi-biometrischen Fusion hervorzuheben.

Acknowledgment

First and foremost I would like to thank my supervisor Prof. Dr. Arjan Kuijper for having the patience and wisdom to guide me throughout this work. His positive leadership and imperative advice have made this journey an exceptional learning experience. I would like to express my gratitude to Prof. Dr. techn. Dieter W. Fellner for co-refereeing this work and for building a thriving atmosphere for scientific excellence at Fraunhofer IGD. Special thanks go to Prof. Dr. Ramachandra Raghavendra for acting as co-referee and for the many fruitful discussions we had over the past few years.

I would like to express my great appreciation to Alexander Nouak and Dr. Andreas Braun. From my first day at Fraunhofer IGD, Alexander has provided me with the full support and trust that always sparked my self-motivation. Andreas constant support and creative leadership gave me the chance to grow scientifically and professionally. I would also like to thank all my friends and colleagues at the Competence Center for Smart Living & Biometric Technology, and previously at the Competence Center for Identification and Biometrics, of the Fraunhofer IGD. Their thoughts and mind-provoking discussions have been an integral part of my life over the past few years. My gratitude extends to the research and non-research staff at Fraunhofer IGD and the Interactive Graphics Systems Group of TU Darmstadt who have taken part in supporting me and have made my work the worthwhile experience that it was. Special thanks are extended to Andreas Braun, Alexander Opel, Aidmar Wainakh, and Philipp Terhörst for their effort in reviewing this thesis.

My greatest thanks go to my students, whom I have learned from the most. Their excellent work, dedication, self involvement, and the many in-depth discussions that we had, made this work possible. Thank you, Philipp, Kristiyan, Yaza, Steffen, Timotheos, Dirk, Hazem, Viola, Christian, Fadi, Benedikt, Chadi, Veronica, Fabian, Aidmar, Tina, Wael, and Matthias.

During my work on this thesis, I worked on a number of European, national, and industrial projects at Fraunhofer IGD. Many thanks go to the project partners at different organizations and institutions. The valuable discussions we had provided me with the clarity and practical relevance much-needed for my research. I am particularly grateful for the many talented minds in the biometrics research community. Their hard work and innovative research have always driven my aspiration, and I am happy to consider many of them friends.

My appreciation is extended to all my friends, across time and space, for the continuous encouragement. Last but not least, I would like to thank my dear family. My brother and sister that I will always look up to, my late father whose wisdom still guides me today, and my mother for her constant and unquestionable support. Looking at my parents personal, academic, and professional lives made this work always feels achievable.

Contents

List of abbreviations

BFW	Brute force weighting
BSSR1	Biometric scores set BSSR1
CMC	Cumulative match characteristic
CNN	Convolutional neural networks
DCT	Discrete cosine transform
DPW	D-Prime weighting
EER	Equal error rate
EERW	Equal error rate weighting
ERM	Empirical risk minimization
FAR	False acceptance rate
FDR	Fisher discriminant ratio
FDRW	Fisher discriminant ratio weighting
FIR	False identification rate
FNR	False negative rate
FPR	False positive rate
FRR	False rejection rate
GBC	General Borda count
GDBC	General distance Borda count
GMM	Gaussian mixture model
HOOF	Histogram of oriented optical flow
HTER	Half-total error rate
ID	Identity
LBP	Local binary pattern
LBLDA	Local binary linear discriminant analysis
LDA	Linear discriminant analysis
LFW	Labeled faces in the wild database
LSH	Locality sensitive hashing

MAD	Median absolute deviation
MEW	Mean-to-extrema weighting
MLDF	Multilobe differential filters
MSU	Michigan state university
NBIS	National institute of standards and technology biometric image software
NCW	Non-confidence width
NCWW	Non-confidence width weighting
NDR	Neighbor distance ratio
NN	Neural network
OLD	Overlap deviation
OLDW	Overlap deviation weighting
OM	Ordinal measures
PAD	Presentation attack detection
PAN	Performance anchored normalization
PCA	Principle component analysis
RBF	Radial basis function
RIR	Rotation invariant representation
ROC	Receiver operating characteristic
SIFT	Scale-invariant feature transform
SRM	Structural risk minimization
SVM	Support vector machines
SVR	Support vector regression
TAR	True acceptance rate
TIR	True identification rate
XM2VTS	Multi Modal Verification for Teleservices and Security applications database
Z-score	Zero mean normalization
TPR	True positive rate

1 Introduction

Biometric recognition is defined as the automated recognition of individuals based on their behavioral or biological characteristics [Int12]. Biometrics were originally used within forensic investigations. However, the application scenarios evolved into other security and convenience aspects. This strong link between individuals and identities is utilized with a focus on security, as in border control and forensic applications, or on convenience, as in automatic log-in and smart-home personalization.

The identity recognition tasks are conventionally achieved by knowledge proofs of identity (e.g. user-names and passwords) and/or physical proofs of identity (e.g. tokens). The knowledge and physical proofs can be easily forgotten, stolen, and imitated. On the other hand, biometric characteristics are developed individually for each person as a fetus by a large extend result of a random process and remain relatively constant for the entire adult lifespan.

In a study provided by Verizon investigators analyzing 47000 security incidents in 2012, 76% of network intrusions studied involved weak credentials [Ver13]. Attacks involving guessing passwords, trying out other passwords related to the same system, and cracking authentication using special tools, made up around four out of five of the hacking breaches in 2012. Verizon estimates that 80% of these breaches could have been avoided if "suitable replacement" to password authentication was used. Biometric authentication is a frontrunner as a candidate for such a replacement.

The advances in the technology and the growth of political and social acceptance resulted in the initiation of much-needed large-scale biometric systems. An obvious example of such projects is the Indian e-Aadhaar project started by the unique identification authority of India (UIDAI) [e-A15]. e-Aadhaar aims at registering the demographic and biometric data of all residents of India, where an estimated 500 million people have no form of reliable identification [Rom11]. e-Aadhaar will grant these people the access to governmental benefits, banking, and the formal economy. A European example of large-scale biometric system is the European union visa information system (VIS) [CotEU04]. The VIS contains biometric information of third country nationals applying for a visa to enter the Schengen area. This allows verifying that the visa holder is the person who applied for it.

This need for more accurate, convenient, and large-scale biometric solutions drove the interest in combining multiple sources of information to form biometric decisions, i.e. multi-biometrics. This work tackles a number of research questions raised by the practical goals of multi-biometrics and offers solutions that help achieve these goals. This chapter will present a more structured motivation for multi-biometric solutions. This is followed by stating the research questions on which this dissertation is based. Finally, an outlook for the next chapters is presented.

1.1 Towards multi-biometrics

In order to discuss the current challenges in the field of applied biometrics, and more specifically in multi-biometrics, the main desired properties of biometric based solution have to be presented. These properties [JBP99] are grouped in seven points:

- *Universality*: a biometric system has to be designed to cover the largest possible ratio of the population. Related problems can include the absence or degraded quality of a certain biometric characteristic in a number of the population.
- *Uniqueness*: a biometric system has to assure to represent different individuals in highly distinct manner.
- *Performance*: the decision errors produced by biometric system is minimized and the computational efficiency is maximized.
- *Permanence*: the performance of a biometric system should be consistent over time. The aging of certain biometric characteristics is a major challenge.
- *Collectability*: the biometric characteristics should be measurable and the quantitative results are reproducible.
- *Acceptability*: the convenience for the user is considered and the usability is maximized.
- *Circumvention*: collecting and replicating a fake biometric sample/template is hard.

Two main and rabidly growing application scenarios are raising the demand on the aforementioned seven properties. The first is the socially and security driven aspect of large-scale biometric systems, such as systems related to identity management, immigration, or criminal investigations. This aspect requires covering numerous number of individuals and be highly accurate. The accuracy especially comes into perspective when discussing open-set identification tasks and duplicate enrollment checks, as the errors grow linearly with the database size. The second demanding use scenario is focused on the aspect of convenience and is driven by individual and industrial users. Such a system can be seen in biometric authentication of smart-devices users, biometric authorized payments, or smart living applications. These solutions have to insure acceptable level of accuracy while maintaining the convenience and usability and adhering to the limitations of device manufacturers. This usually results in a sub-optimal capture conditions and thus pressure the performance requirements.

Biometric systems in different applications conventionally used one biometric information source (characteristic, algorithm, instance, or capture) to verify or identify a subject identity. This has limited the accuracy of such systems to the cap accuracy of the considered source. Pushing biometric systems to the border of their *performance* triggered the use of multiple biometric information sources within a unified decision-making process aiming at more accurate, universal, convenient, and secure systems.

Using multiple biometric sources increases the *universality* by offering more options for the users, e.g. a user with a low quality fingerprint (unusable) can still use a multi-biometric system that also considers the iris. *Uniqueness* is also improved by multi-biometrics in the cases where multiple individuals share a very similar biometric characteristic but are unique in another.

Having more information in hand, from multiple biometric sources, enables creating more accurate decisions and thus achieving higher *Performance*. The higher accuracy affect other desired properties such as *permanence*, where a number of biometric sources with varying ageing problems may combine into an accurate single decision.

Constructing a multi-biometric system with higher accuracy will allow to ease the restrictions on capture conditions compared to conventional biometric systems, this will result in a higher *collectability* and *acceptability*. Attacking a biometric system that uses multiple biometric sources would require increased effort and knowledge from the attacker and thus would make the *circumvention* harder.

Most of the discussed properties are enhanced inherently by the concept of multi-biometrics. This leaves the focus of research on enhancing the performance of such a system and tackling the functionality challenges. The research efforts will be categorized in this work based on the point of effect in the multi-biometric system structure. The main points of focus are the pre-fusion operations, the optimization of the fusion process by source weighting and the integration of supplementary information, multi-biometric data retrieval, and discussing a number of processes that complement and utilize the multi-biometric fusion process.

1.2 Research questions

Based on the current state-of-the-art, this dissertation tries to enhance multi-biometric technologies through answering a set of open research questions. These questions aim at improving the overall performance and functionality of multi-biometric systems. In order to put these questions in a broader perspective, they will be grouped based on their point of effect in the multi-biometric decision-making process. The first group will focus on the operations performed in preparation to the actual fusion process. The second group will deal with the optimization of the fusion process to achieve higher biometric accuracy. The third is concerned with biometric reference databases in large-scale biometric systems. Finally, the last group covers specific applications that complement biometric systems and utilize multi-biometric fusion to enhance their performance.

To illustrate the relation of the research questions to the mutli-biometric system processes, Figure 1.1 presents a standard multi-biometric system work-flow. This aims at providing the basic understanding of such a system and enabling a structured reasoning of the research questions. As shown in the figure, a number of biometric characteristics are captured then checked for quality compliance, as well as for the possibilities of being a presentation attack on the system, rather than a bonafide presentation. The samples that pass this check are processed to extract representative features (templates), and in some cases, a number of templates are fused to create a more informative and stable one. At this point, the resulting templates are either passed to the reference database in case of user enrollment, or further processed in the biometric system for verification or identification. In case of a verification or identification scenario, these templates are compared to one or more of the reference templates, resulting in a set of comparison scores. These scores are normalized into a common space and any missing scores are accounted for and further processed for completeness. The resulting complete set of normalized scores is optimally fused to achieve the best possible accuracy. The fused scores can be complemented by further information resulting from the presentation attack detection, the quality assessment, or information from the comparison scores, before or after normalization. The result of the fusion process is further processed to make the verification or identification decision of the multi-biometric system.

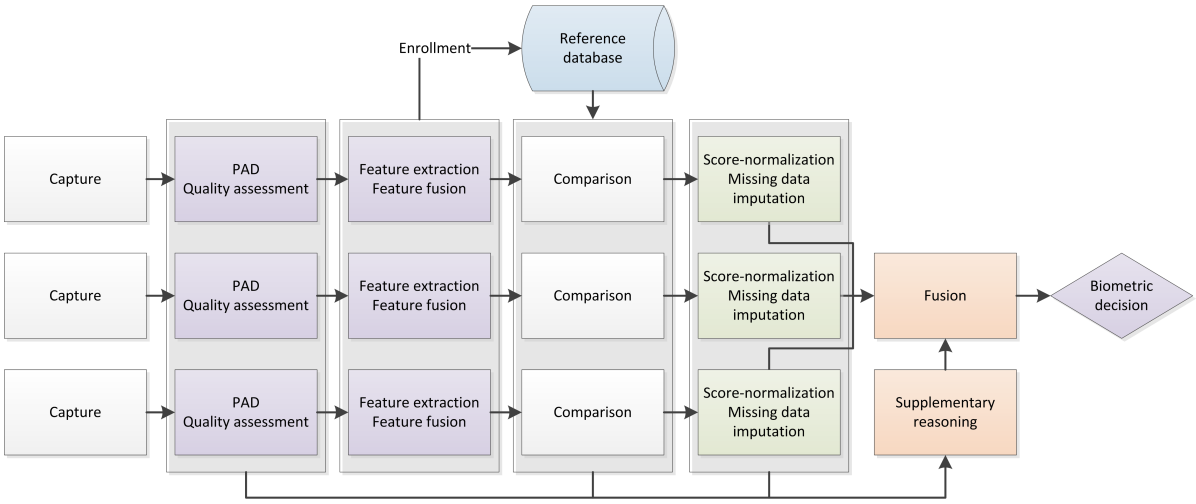


Figure 1.1: typical work-flow of a multi-biometric system. The four research question groups are in different colors, the pre-fusion processes in green, the fusion optimization in orange, the reference database in blue, and the miscellaneous multi-biometric processes in purple.

1.2.1 Pre-fusion processes

Multi-biometric systems integrate information provided by different sources in the fusion process. These sources might depend on different biometric characteristics, capturing sensors, feature extraction methods, or comparators. Therefore, each of these sources will result in similarity scores in a different space of reasoning. A well-designed solution to unify the reasoning space of these scores largely affect the performance of the whole fusion system.

Typically, each biometric comparison is expected to produce a comparison score, describing the similarity between the feature representation of the captured characteristic and a reference representation. Due to failures to pass the comparison or any earlier step, some expected similarity scores might be missing. For a fusion approach that expects a full set of comparison scores, a proper solution to deal with such cases is essential to maintain the functionality while achieving the highest possible performance.

Keeping this in mind, the following research questions are drawn to tackle these issues in information preparation for the fusion process.

- **Performance aligned normalization:** to enable a clear and predictable score-level fusion process, the fused comparison scores have to be brought into a comparable range, this is performed by score normalization. Score normalization techniques heavily affect the performance of multi-biometric systems [JNR05] and thus are an active research field within the study of multi-biometrics. Ideally, a normalized score value has to point out a certain decision in a comparable confidence level regardless of the score source. Jain et al. presented a comparison between different score normalization methods for multi-biometric fusion [JNR05]. However, normalization approaches used in multi-biometric fusion aim only at unifying the range of the scores and, in some cases, their distribution properties. Achieving score normalization that is related to the biometric decision inference of a score value will require building a connection between the operational properties of a biometric source and the normalization process. This leads to the first research question tackled in this work and it can be stated as follows

  > "*RQ1: how can an operation-related score normalization approach be designed and what would be its effect on the overall biometric performance?*".

- **Missing data imputation:** a perfect scenario where all the fused biometric sources provide their comparison scores for the fusion process is not always the case. In some cases, the failure to capture a biometric characteristic with the required quality or the failure to extract representative features from the capture can lead to missing part of the fused scores. Previous works dealt with the problem of missing values and their imputation, although, most of the works considered only the verification scenario [FKP08, Lit92, PWM*10]. These approaches had different levels of complexity without a consistent advantage of complex approaches. The different nature of the identification process may result in different performance reactions for different imputation methods. This motivates one of the driving questions in this work

  > "*RQ2: what would be the effect of different data imputation methods on the identification performance in comparison to the verification scenario? and how much gain can a complex imputation method bring?*".

In an effort to answer these questions, this work presents a performance anchored normalization scheme that aims at aligning the interpretation of the scores provided by different biometric sources. A study of different miss-

ing scores imputation approaches and their relative effect into the verification and identification performances is also discussed.

1.2.2 Fusion optimization

As shown in Figure 1.1, a number of scores results from the different comparisons between the captured biometric characteristics and the enrolled references. These scores are fed into the fusion process where they should be optimally joined to produce an accurate biometric decision. The fusion of these scores is commonly controlled by assigning a relative measure, i.e. weights, for the different sources to control their individual contribution to the fusion result. Therefore, assigning these weights has a strong influence on the multi-biometric system performance. The fusion process also includes supplementary reasoning in the final decision in an effort to enhance the overall accuracy and robustness to sub-optimal capture conditions or environment. This supplementary information can be captured from an early stage in the process, such as detecting presentation attacks or estimating the capture quality. However, such systems are not always reliable and this early access to raw data may not always be granted to the system integrators. Therefore, the information gathered at post-comparison stages can be valuable to supplement the comparison scores and achieve more accurate biometric decisions.

The questions listed in the following aim at further optimize the fusion process to enhance the overall biometric performance.

- **Multi-biometric source weighting:** simple information fusion approaches such as the sum rule score-level fusion proved to achieve high performance compared to more sophisticated approaches [RJ03]. An optimization of that would be the weighted-sum rule, where each biometric source is weighted to indicate its relative importance, and thus contribution, to the final fused biometric decision.

  Weighting based on the equal error rate of biometric sources is widely used [JNR05] along with approaches based on D-Prime calculations [SUM*05] and Fisher discriminant ratio [PB04]. However, these weights might not capture the different properties that make a biometric source more reliable than another, therefore, they might not achieve dependable generalization. This opens the next question

  > "*RQ3: what weight can be derived, so that it captures the absolute performance of a biometric source as well as its confidence?*".

  A number of previously proposed multi-biometric source weighting methods were based on verification performance metrics, such as the equal error rate [JNR05] and the properties of the receiver operating characteristics [TKL08, VP09]. These weights are then used generically for both verification and identification tasks. This raises the next question

  > "*RQ4: what could be good multi-biometric source weighting candidates based on identification performance? Would such a weight positively affect the multi-biometric identification performance?*".

- **Supplementary information in multi-biometric fusion:** conventionally, score-level multi-biometric fusion exclusively uses the biometric comparison scores provided by the fused biometric sources and general information about these sources (e.g. weights). Previous works extended this concept to include sample quality information related to each biometric comparison [NCJD06, PK08, PMK09]. One source of possible supplementary information is the relative relation between different comparisons, which leads to the next question

> "*RQ5: what discriminant information can be derived from the relative relations between different comparisons? Does integrating this information enhances the multi-biometric performance?*".

Also on the score-level, the relation between the scores from different sources in the same multi-biometric comparison, rather than different comparisons, can contain rich information leading to a better control over the fusion process. This view leads to the question

> "*RQ6: what discriminant information can be derived from the relation between the scores of different sources in a multi-biometric comparison? Does integrating this information enhances the multi-biometric performance? Does this make the multi-biometric system more robust to the realistic scenario where noisy captures occur?*".

Focusing on the weighted-sum score-level fusion, this work presents a weighting approach that aims at representing both, the accuracy and the confidence of the fused biometric sources. Under the same scheme, weights based on the properties of the cumulative match characteristic curve were suggested and their effect on the identification performance was discussed. Supplementary information in the form of neighbor distance ratios and score coherence measures were suggested and introduced into the fusion process in this work.

1.2.3 Reference database

Whether under the verification or identification scenario, a captured biometric characteristic is compared with one or more reference templates to measure their similarity. The reference templates are stored during enrollment in a reference database. Performing a 1:N identification search in such a database can require a huge computational effort and time. Indexing structure can help limit the space of these searches, however with some accuracy loss. The availability of multi-biometric references can help enhance the performance of reference database indexing. Aiming at enhancing the search speed while maintaining high accuracy by utilizing multi-biometric sources, the following research question is stated.

- **Multi-biometric data retrieval:** previous multi-biometric indexing solutions focused either on the feature or the rank-level fusion. The feature-level is limited by the possible combinations of modalities and algorithms used by each biometric source. Moreover, it might face challenges in situations where some sources are missing. Solutions focusing on the rank-level are more flexible in terms of different biometric sources and missing data. However, these solutions disregard in-depth information and treat every rank and every biometric source equally [GR09, GR12]. Moreover, errors in indexing-based multi-biometric data retrieval are usually caused by the cases where single source candidate lists are of low quality. However, no previous work targeted or analyzed these situations. The pursue of a flexible and optimized multi-biometric indexing solution leads to the following question

  > "*RQ7: how to design a multi-biometric data retrieval structure that benefits of the rank-level fusion flexibility while still being optimizable to different sources? What supplementary information can be added to enhance the indexing performance? How would this affect the critical cases where single sources supply low quality candidate lists?*".

With a focus on the flexibility of rank-level multi-biometric data retrieval, this work presents a solution that can be adapted to biometric sources of different nature. By proposing the use of efficiently calculated index-based approximated distances, more information is fed into the retrieval process. This aims at maintaining low computational cost while enhancing the accuracy of the retrieval, especially when fusing candidate lists of low quality.

1.2.4 Miscellaneous multi-biometric processes

Besides conventional biometric verification and identification tasks, multi-biometric fusion is used within different biometric processes. These processes complement the multi-biometric framework seen in Figure 1.1. An example of such a process is the presentation attack detection used to verify the validity of the capture. Another example is the feature-level fusion that can be used to fuse a number of biometric templates into one. Such a process can also change the application perspective of the multi-biometric fusion, an asynchronous fusion can allow fusing different types of biometric sources to achieve a continuous authentication over time.

- **Multi-biometric continuous authentication:** biometric authentication is typically used to gain access (e.g. log-in) to a system/service. This means that the individual is recognized once at the start of a process that might span over a period of time. However, an attacker could gain access to the system after this initial log-in. Continuous authentication can be used to protect from such attacks. Continuous authentication also introduces some constraints to a typical biometric system as it should collect biometric samples continuously and conveniently (no user collaboration). Previous works focused on using behavioral biometric characteristics such as keystroke dynamics [BW12, ZD15, BM15]. Using multi-biometrics, including biological characteristics, might help enhance the convenience and security of such systems. The question here is

  > "*RQ8: how is such a system designed? What would be its effects on the performance compared to single source solutions? What would be the main challenges in developing and deploying such a system in realistic scenarios?*".

- **Face presentation attack detection:** the recent availability of very accurate and efficient face recognition algorithms leaves the vulnerability to presentation attacks as the major challenge to face recognition solutions. Previous works showed good performances on research databases. However, face presentation attacks are still an open problem in practical scenarios [WHJ15]. Biometric fusion might help enhance the performance of presentation attack detection. However, some open questions are still to be answered, such as

  > "*RQ9: how well a presentation attack detection performs under realistic conditions? What is the effect of feature and score-level fusion on the performance? and how long (time) should the face capture last to saturate the fusion performance?*".

- **Face selection and template fusion:** face recognition from videos can be used for surveillance and convenience applications. Creating a face reference from a video capture can benefit from the multiple images available in a video, and thus create a more informative representation. Having a single reference representing a video sequence enhances the computational efficiency of such a system. However, some questions regarding this process are left unanswered, and they can be stated as

> "*RQ10: how should faces be selected from a video sequence? What is the best approach to fuse their features into a unified template? and would this outperform simple, but not computationally efficient, score-level fusion?*".

Getting beyond the standard recognition role of multi-biometrics, different biometric processes are presented in this work that utilized and complimented muti-biometric fusion to achieve their goals. Later in this dissertation, applications focused on continuous authentication, face presentation attack detection, and face recognition in videos are discussed.

1.3 This thesis

After motivating and introducing the research focus of this work, the rest of the thesis will be organized as follows:

Chapter 2 presents the essential background information regarding the typical structure of an automated biometric system. A more detailed look into multi-biometric systems and information fusion is discussed along with a high-level view over the related works in the literature. The biometric performance metrics commonly used in the literature are also presented.

Chapter 3 discusses two issues of pre-fusion processes in multi-biometric score-level fusion. These issues are the performance related normalization and the missing data imputation. First, as a response to *RQ1*, this chapter presents a novel normalization technique that tries to align the inference of different score sources by aligning a certain performance-related point in the score distributions. The performance anchored normalization (PAN) algorithms proposed in this chapter are evaluated along conventional score normalization techniques and used combination fusion rules to produce the fused scores. Chapter 3 will then focus on missing data imputation within score-level multi-biometric fusion. Different solutions are presented and their effect on the overall performance is compared between the identification and verification scenarios. A solution based on support vector regression was proposed and compared to simpler approaches. The work on missing data imputation aimed at answering RQ2 stated in Section 1.2.1.

Chapter 4 proposes a number of biometric source weighting approaches and evaluates them along with the state-of-the-art and best-practice techniques. These weights are utilized to optimize the effect of different biometric sources on the final biometric decision. Within this prospect, two paths are followed as an answer to *RQ3* and *RQ4*. First that considers the properties of the genuine and imposter scores distributions to capture the overall performance of the biometric source, as well as an indication of its confidence. This solution aims at avoiding the shortcomings of previously proposed solutions such as low generalization abilities and sensitiveness to outliers. The second path focuses on deriving the biometric source weights from an identification performance representation, the cumulative match characteristic curve. Different features of this curve are evaluated as candidates for being optimal weights in a weighted-sum fusion biometric approach.

Chapter 5 is concerned with introducing supplementary information into the multi-biometric fusion process to further optimize its decisions. Beside comparison scores from multiple biometric sources, this chapter proposes another evidence on the genuinity of a comparison. This evidence is based on the relation of the fused scores to other comparisons scores. An initial approach demonstrates the effect of integrating this supplementary information on the performance. A more optimized solution is also presented and it includes taking advantage of biometric source weighting (discussed in Chapter 4) in an initial fusion process. This provides answers to the previously stated *RQ5*. This chapter also presents supplementary information based on the relative relation be-

tween the scores in one multi-biometric comparison. This relation is modelled as coherence information between these scores and integrated into the fusion process as a response to *RQ6*.

Chapter 6 presents a generalized solution for multi-biometric data retrieval. This can include indexing structures of different nature in an optimized manner. The presented approach generalizes Borda count approach by adapting it to different biometric indexing sources. This is also extended to include efficiently calculated approximated distances based on each source indexes. This is evaluated over combinations of single-source candidate lists of different qualities and proved to largely enhance the retrieval performance when faced by low quality candidate lists. The proposed approach is evaluated under the multi-instance iris identification scenario and compared to a number of state-of-the-art and best-practice approaches. Chapter 6 responds to *RQ7* by presenting, discussing, and evaluating this solution.

Chapter 7 focuses on different application where multi-biometric fusion is used to complement the conventional verification/identification scenario. Three applications are discussed where the multi-biometric fusion plays a major role in enhancing the overall performance. First application aims at using behavioral and physical biometric characteristics to assure a continuous author authentication. In this scope, a database was collected and an asynchronous fusion approach is presented, evaluated, and discussed as a response to *RQ8*. The second application deals with the problem of face presentation attack detection. The practical use of presentation attack detection is discussed by investigating the more realistic scenario of cross-database evaluation and presenting a state-of-the-art performance comparison. The relation between the video duration and the detection performance is also investigated. This is done along with presenting an optical flow based approach that benefited from different multi-biometric fusion approaches. The second application is directly related to *RQ9*. Chapter 7, as a response to *RQ10*, then discusses utilizing multi-biometric fusion to create face references from videos. Face selection, feature-level fusion, and score-level fusion solutions were evaluated under the scenario of face recognition in videos.

Chapter 8 concludes this work by highlighting its contributions, elaborating its practical benefits, and giving an outlook for future research.

Figure 1.2 links the technical chapters of this work and the research questions they answer to the components of the multi-biometric work-flow, previously presented in Figure 1.1.

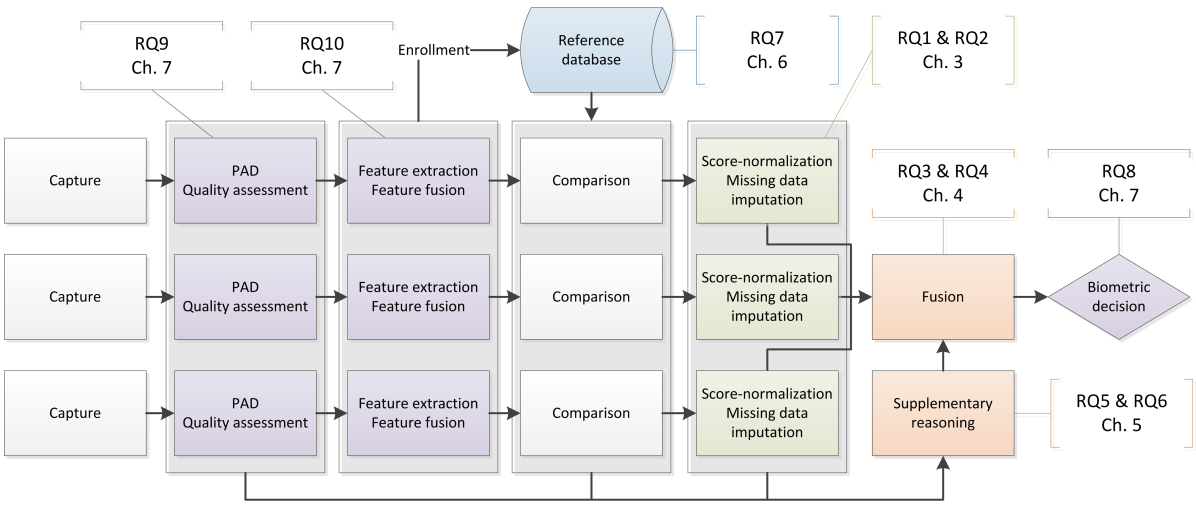


Figure 1.2: the chapters and research questions as part of the multi-biometric work-flow.

1.4 Conclusion

This introduction provided a motivation leading to a set of research questions that established a structure for the contributions in the rest of the thesis. The listed questions aimed at advancing the state-of-the-art, related to different blocks of the multi-biometric system work-flow, in an effort to enhance the overall performance and usability. These advances were grouped in four categories namely, the pre-fusion processes, fusion optimization by source weighting and integrating supplementary information, multi-biometric reference data retrieval, and miscellaneous multi-biometric processes that complement and utilize multi-biometric fusion.

The research questions within the pre-fusion processes focused on the validity of building a link between score normalization and its inference on the biometric performance. It also targeted the issue of missing data imputation within score-level multi-biometric fusion. The second group of research questions focused on optimizing the score-level fusion operation by motivating weighting approaches that consider single source confidence as well as investigating identification performance metrics behavior as fusion weights. Optimizing the fusion process drove a second set of questions that are concerned with introducing supplementary information derived from the relation between comparison scores in different multi-biometric comparisons and within the same multi-biometric comparison set.

The third focus of the presented research questions was concerned with the multi-biometric data retrieval by stating the need for an indexing structure that takes advantage of the flexibility of rank-level fusion, while maintaining essential information available on the feature-level. The fourth research questions group targeted processes at the edges of the multi-biometric system work-flow, the processes that utilize and complement conventional multi-biometric systems. Three questions were stated about multi-biometric continuous authentication, face presentation attack detection, and creating face reference templates from video sequences.

2 Background

The previous chapter presented a motivation and a structure for the research problems dealt with in this thesis. To facilitate a better understanding of the following chapters, this chapter presents the typical properties of an automated biometric solution. This includes a more detailed insight into multi-biometric fusion while presenting a high-level discussion of the related works in the literature. Biometric performance metrics commonly used in the literature are also discussed. This chapter is partially based the published work [DOS13].

2.1 Biometrics

Individual characteristics are used by people in their daily life as the mean of intra-human recognition. Overtime, these discriminant characteristics found an application in forensic subject identification. This initially started in the late 19th century by using a set of body measurements by Alphonse Bertillon of the Paris police department [Rho56]. By the beginning of the 20th century, fingerprint analysis has become a main tool in solving many crimes [Hue09] .

Modern biometric recognition, as seen from a computer science perspective, is defined as the automated recognition of individuals based on their behavioral or biological characteristics [Int12]. In order to choose a certain characteristic, it has to be evaluated in terms of the desired properties. These properties [JBP99], as mentioned in Chapter 1 are:

- *Universality*: a biometric system has to be designed so that it is able to cover the largest possible ratio of the population. Related problems can include the absence or degraded quality of a certain biometric characteristic in a number of the population.
- *Uniqueness*: a biometric system has to assure to represent different individuals in highly distinct manner.
- *Performance*: the decision errors produced by biometric system is minimized and the computational efficiency is maximized.
- *Permanence*: the performance of a biometric system should be consistent over time. The ageing of certain biometric characteristics is a major challenge.
- *Collectability*: the biometric characteristics should be measurable and the quantitative results are reproducible.
- *Acceptability*: the convenience for the user is considered and the usability is maximized.
- *Circumvention*: collecting and replicating a fake biometric sample/template is hard.

The mentioned properties are associated differently with different possible biometric characteristics. This was previously discussed [JBP99] and is presented in Table 2.1. For example, face biometrics are highly acceptable and measurable, but it is less accurate than other characteristics like fingerprint and iris. Fingerprint, on the other hand, suffers from relatively lower universality because of the low fingerprint quality that might be a result of erosion by intense physical contact. Iris is less acceptable and measurable as, until recently, it required user collaboration to perform infrared scans.

Characteristic	Universality	Uniqueness	Permanence	Measurability	Performance	Acceptability	Circumvention
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vein	M	M	M	M	M	M	L
Iris	H	H	H	M	H	L	L
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Voice	M	L	L	M	L	H	H

Table 2.1: comparison of biometric characteristics (H: high, M: medium, L: low) [JBP99].

A biometric sensor captures the biometric characteristics. In a typical biometric work-flow, a set of features are then extracted from each captured sample. This feature set (template) is either stored in a database as a reference in the enrollment process or compared to a previously acquired biometric feature set (reference template) stored in a database. A comparison results in a similarity score. Analyzing the similarity score (or a set scores) results in a biometric decision.

A biometric decision, and thus a biometric system, can be a verification one or an identification one. Biometric verification is the use of biometrics information to verify a person claimed identity. The identity can be claimed using a smart card, a user name, or an identification number. Here, the system will verify that the claimed identity belongs to the user by comparing his/her biometric characteristics with a stored (and associated to this identity) biometric template. Therefore, the comparison (similarity measure) is only performed once for each identity claim and thus, the verification problem is usually referred to as a $1:1$ comparison process.

Biometric identification tries to assign an identity to an unknown person based on the captured biometric characteristics. This requires comparing the captured biometric characteristic to all the enrolled subjects. Therefore, the identification is referred to as a $1:N$ comparison process, where N is the number of the enrolled subjects.

Biometric identification can be categorized into open-set and close-set identification. Close-set identification system is confident that the captured subject is one of the enrolled subjects, and thus can report the best matching identity as the identified one. Open-set identification implies that the system does not guarantee that the captured subject is already enrolled. Therefore, the open-set identification system has to verify one of the top matched identities to be the identity of the captured subject or to point out the fact that the subject is not enrolled.

To include a subject in a database, an enrollment process has to take place. Enrollment includes providing a trusted identity, capturing and quality proofing a biometric characteristic, feature extraction, and finally storing

the identity information and the extracted features (template) in a database. Figure 2.1 illustrates the three typical operations of a biometric system, enrollment, verification, and identification.

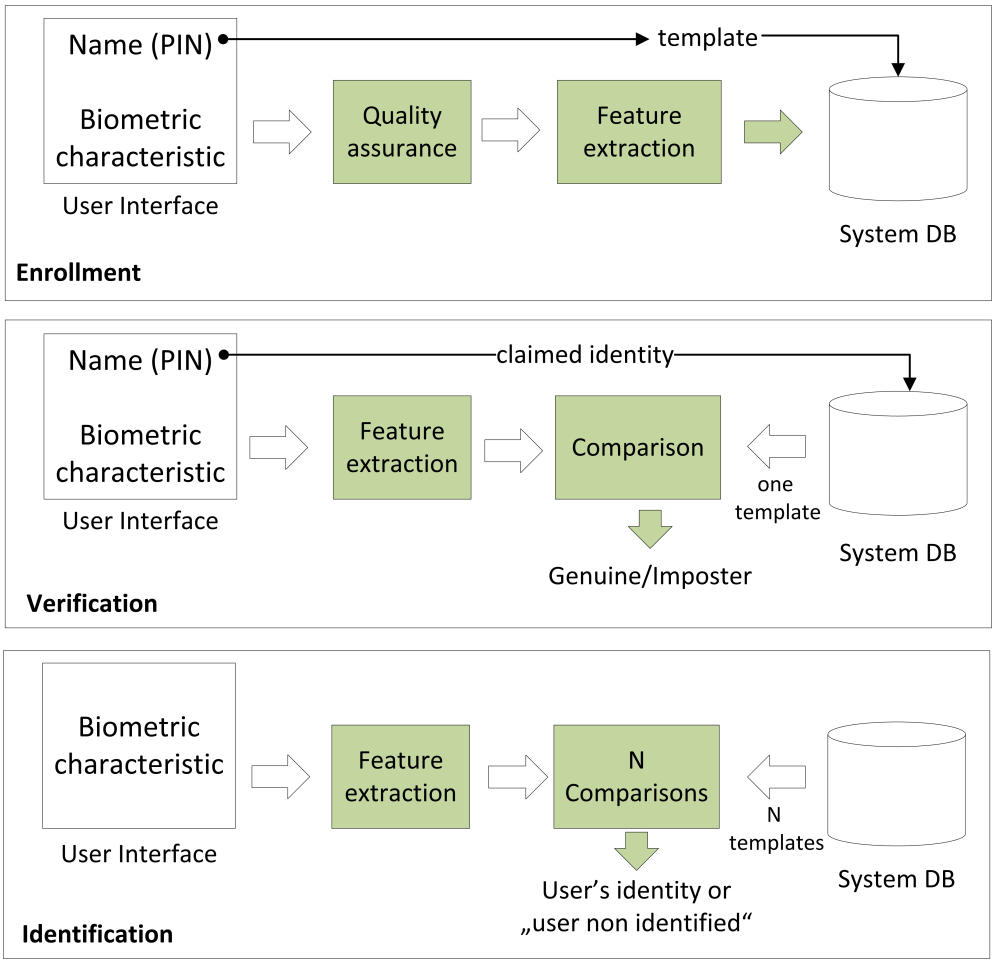


Figure 2.1: the main operations of a biometric system: enrollment, verification, and identification [JRP04].

2.2 Multi-biometric fusion

2.2.1 Information fusion

Information fusion is the integration of multiple information concerning a certain aspect to produce a more accurate description or decision. Information fusion aims at building more accurate decisions (descriptions) by obtaining synergy [Mit12]. Synergy here points out that a grouped representation of an object is better and more complete than an individual component representation. According to Bellot et al. [BBC02], the effect of information fusion can be summarized in the following points:

- Accuracy gain: decisions and representations obtained after the fusion process are more accurate. Noise and errors are reduced in comparison to single source information.
- Completeness gain: the information represented by fusion is more complete and less redundant.
- Representation gain: fused information should be more granular compared to each of the initially fused information.
- Certainty gain: fusion induces an increase of the belief in information after the fusion process.

This set of gains can be utilized to improve the desired characteristics of a biometric system.

2.2.2 Multi-biometrics

Multi-biometrics uses more than one biometric information source in a unified frame in an effort to solve problems faced by the conventional uni-modal biometrics. The multi-biometric approach aims at improving biometrics by increasing accuracy, and robustness to intra-person variations and to noisy data. It also intends to solve uni-modal biometrics problems with universality and vulnerability to spoof attacks.

Information fusion in multi-biometrics is used to build an identification/verification decision based on the information collected from different biometric sources. The fusion can be done on different levels, which can be seen as a tradeoff between information loss and integrability. Data and feature-level fusion maintain a large proportion of the original information but force limitations on the design and implementation. Only a limited combination of information sources can be fused and building a system out of different hardware components might be impossible. On the other hand, decision-level fusion can be seen as naive and lacking the opportunity to use important information. Based on this, the score-level fusion makes a good tradeoff between integrability and maintaining important decision information. Score here refers to the comparison score (similarity) between each biometric capture and a stored reference. Figure 2.2 presents an overview of multi-biometric systems with different levels of fusion. The different biometric information sources can be based on different characteristics, algorithms, instances, sensors, or presentations.

Figure 2.3 presents an overview of multi-biometric identification system using score-level fusion. Scores are produced by comparing captured characteristics to stored reference ones. The resulting scores from different sources (algorithms and modalities) are normalized into a comparable domain, then passed into a fusion algorithm. The fusion then results in fused scores that a verification/identification decision can be built upon by thresholding and/or ranking.

In the following, topics covering different stages of the multi-biometric fusion process are discussed and connected to the literature. This section aims at serving as a general description of multi-biometric fusion while presenting a high-level discussion of related works. Deeper insight into related work concerning the specific contributions of this dissertation will be presented in the respective chapters.

Biometric sources: multi-biometric systems are categorized depending on the nature of the fused biometric sources from which a unified final decision is built. The main approaches are multi-modalities, multi-algorithmic, multi-instance, multi-sensorial, and multi-presentation.

Multi-modalities is the use of more than one biometric characteristic as an identity measure. Some works combined fingerprints and face images [KNR09, TWL10, KTT10], others fused fingerprints and iris biometrics [BBKQ09]. Using face images along with iris biometrics was also introduced [WTJ03]. One of the most interesting multi-modal approaches is the use of ear and face biometrics, as they can be easily and non-intrusively captured using same or similar devices [CBSV03, Yan06]. Other combinations were also introduced, such as ear and fingerprint biometrics [RKBT06].

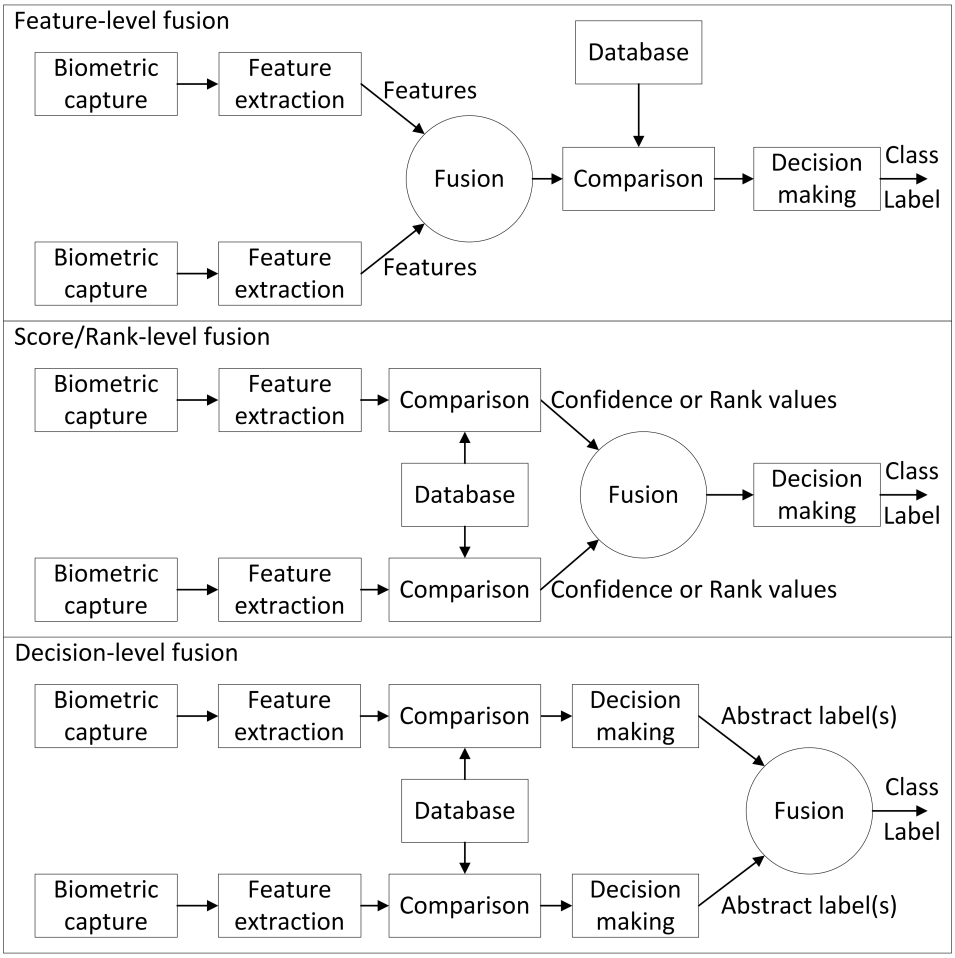


Figure 2.2: multi-biometric fusion at different levels [JRP04].

Many works dealt with multi-algorithmic biometrics, such as using multiple face matchers [KNR09, BKTR10, KTT10], or ear identifiers [MSV99, YB05]. Multi-instance fusion was also studied, such as two different fingerprints [YB05, BKTR10, KTT10], multi-sensorial [AH06], and multi-presentation biometric fusion [CTG11] were also discussed in the literature. However, the nature of the fused biometric sources is usually limited by the application scenario and goals.

Score normalization: the scores to be processed by the score-level fusion algorithms are usually not homogeneous as they are produced by different sources. These scores have to be brought into a common comparable range by a normalization process. Some of the most common normalization techniques are min-max normalization, z-score normalization, double sigmoid, TanH-estimator, and median absolute normalization. The parameters that rule the normalization process are determined based on the statistical properties of the training data. The performances of normalization techniques are not directly comparable as they depend on the overall multi-biometric system.

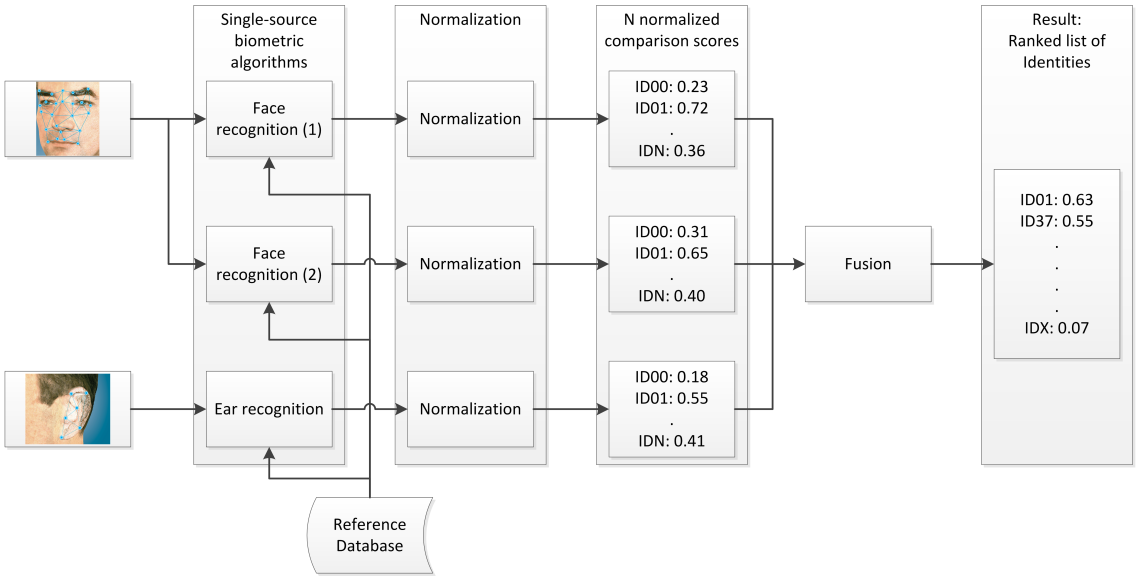


Figure 2.3: a high level structure of a multi-biometric identification system using score-level fusion. Comparison results are produced by different algorithms and modalities then normalized. The sets of normalized scores feed the fusion algorithm to produce a final set ranked by fused comparison scores. The algorithms block encapsulates quality assessment, feature extraction, and template comparison.

The min-max normalization depends on the range that the scores span regardless of the distribution properties and aims to map the scores into a range of [0,1]. This normalization scheme was used successfully by many works dealing with score-level multi-biometric fusion [NS09, VIM*07]. The min-max normalization scheme is highly sensitive to outliers [JNR05] as it depends on single minimum and maximum values. Z-score normalization is a more sophisticated score normalization method and was used in several works [NS09, VIM*07]. Here, the arithmetic mean (μ) and the standard deviation (σ) of the score values are considered. This method assumes a Gaussian distribution of the score values. This normalization method has low robustness as well, as the parameters μ and σ are sensitive to outliers.

The median absolute deviation normalization method is similar to the z-score normalization method but uses the median and median absolute deviation instead of the mean and standard deviation. This method is more robust to outliers but it also assumes a near Gaussian distribution of the comparison score values.

The double sigmoid function is another normalization method. This normalization maps the scores into a range of [0,1] and requires fine-tuning of its parameters. The double sigmoid normalization was used to combine finger print comparison scores by Cappelli et al. [CMM00]. Marsico et al. recently proposed a novel normalization approach named the Quasi-Linear Sigmoid (QLS) [DNRT11]. This approach aimed at overcoming the sensitiveness of the traditional normalization algorithms to outliers.

The selection of a proper normalization method is a tradeoff between efficiency and robustness, and it depends largely on the nature of the application. Methods like min-max normalization and z-score normalization tend to be more efficient. On the other hand, median absolute deviation normalization and double sigmoid function normalization are usually more robust but require higher computational effort [JNR05].

Score-level fusion: score-level fusion algorithms can be categorized into two main types, combination rules and classification based fusion. Combination rules are simple operations performed on the normalized scores. These rules produce a combined score, and a classification decision is made based on this combined score value. Main combination rules are the sum rule, weighted-sum rule, product rule, max rule, min rule and median rule.

Some works discussed the difference in performance between the combination rules. Most studies showed the superiority of sum and product combination rules [HS12, NCJD06, CBFC04]. One must keep in mind that different combination of sum rules and normalization techniques produce different results [LL11].

Classification based fusion considers the input score values as a feature vector. Given these vectors, a classifier is trained to classify a new given vector into genuine or imposter class. Different types of classifiers can be used, just as neural networks [Als10], k-nearest neighbors (K-NN) [JSH04], support vector machines (SVM) [SVN07, GMAD05], Adaboost [IKWY10, MP09], or as likelihood ratio [NCDJ08, IR10] classifiers. Some works showed comparable results between combination rules and classification based fusion [RCLM08, MVSM12]. Other works showed the superiority of combination rules [SVN07].

However, dealing with a small number of inputs using complex classification-based fusion techniques, can only estimates separation lines optimal for the separation of overlapping classes while having small degree of liberty [RDRK11]. Simpler fusion approaches, such as weighted-sum rule, proved to achieve comparable performances to more complex approaches [NCDJ08, RNJ06] while maintaining simpler and easily integrable design.

Supplementary quality information: in order to improve accuracy and compensate for missing information when moving from feature-level fusion to score-level fusion, the quality of the biometric samples is considered as *supplementary information* in the fusion process.

The quality of the captured biometric sample (image/scan/recording) has an effect on the comparison score values and the confidence of these values as the features extracted from these samples are not reliable [NCJD06]. This reflects on the role that a certain score value plays in the fusion process, and therefore, taking the quality measures into account will enhance the performance of the fused multi-biometric system [NCJD06, PK08, PMK09].

Nandakumar et al. [NCJD06] proposed a fusion algorithm that takes into account the sample quality into their likelihood ratio-based fusion scheme. Their experiments proved a highly positive impact of considering the quality measures. Poh et al. [PMK09] proposed a classifier based fusion algorithm that considers the biometric sample quality, as well as, the biometric capture device information. Their experiments clearly showed the effect of including the quality measures on performance.

Missing data: in practice, some comparison scores can be *missing* because of a missing modality or a low quality capture. To build a robust multi-biometric system, the possibility of missing score values must be considered and dealt with so that a reliable biometric decision is made.

Many works considered the problem of missing data and proposed solutions for robust fusion algorithms [PWM*10, KNR09, DDV07]. However, most of these works dealt explicitly with the fusion problem under the verification scenario. Nandakumar et al. proposed a robust fusion solution for multi-biometric fusion under the identification scenario that aims to produce an identification decision regardless of the partially missing data [KNR09]. The authors extended the likelihood ratio-based score fusion (originally designed for verification problems) to perform under the identification scenario.

2.3 Performance metrics

While most biometric characteristics are theoretically discriminant, automatic biometric systems make wrong verification or identification decisions in some cases. To build a performance comparison between different biometric systems, a set of performance metrics are defined for different operational scenarios. This section presents the relevant metrics and is largely based on the international standard ISO/IEC 2382-37:2012 [Int12].

Some of the errors in biometric systems occur in an early stages of their operation, namely in the capture or feature extraction stages. These errors are:

- Failure to capture rate (FTC): proportion of failures of the biometric capture process to produce a captured biometric sample that is acceptable for use.
- Failure to extract rate (FTX): proportion of failures of the feature extraction process to generate a template from the captures sample.
- Failure to acquire rate (FTA): proportion of verification or identification attempts for which the system fails to capture or locate an image or signal of sufficient quality.

FTA is a result of FTC or FTX and can be given as:

$$FTA = FTC + FTX * (1 - FTC). \tag{2.1}$$

The FTC, FTX, and FTA are related to the performance of the complete biometric system and thus are usually neglected when measuring the performance of the biometric algorithms.

Performance metrics of the biometric algorithms can be categorized depending on the operational mode, verification or identification.

2.3.1 Verification performance metrics

Biometric verification builds its binary decision (genuine/imposter) by thresholding the level of similarity between the probe and the reference samples (comparison score). Assigning a threshold value is a tradeoff between security and convenience. Security requires a low rate of imposter users accepted as genuine (false positives) and convenience requires a low rate of genuine users rejected as imposters (false negatives). This tradeoff can be summarized by the following biometric verification metric:

- False match rate (FMR): rate of zero-effort imposter attempt samples falsely verified as genuine (match) to the compared sample.
- False non-match rate (FNMR): rate of zero-effort genuine attempt samples falsely verified as imposter (non-match) to the compared sample.

The genuine match rate (GMR) is another metric used in the literature and points out to the same properties as the FNMR. The GMR is given by

$$GMR = 1 - FNMR. \tag{2.2}$$

The discussed FMR, FNMR, and GMR are error rates that describe the algorithm verification performance. Therefore, they do not consider errors introduced by the whole system, i.e. FTA. A set of similar verification metrics that theoretically consider FTA errors is also used in the literature, these metrics are the false acceptance rate (FAR) and false rejection rate (FRR) and are given by

$$FAR = FMR * (1 - FTA) \tag{2.3}$$

and

$$FRR = FTA + FNMR * (1 - FTA). \tag{2.4}$$

The true acceptance rate (TAR) is derived from the FRR and is given as

$$TAR = 1 - FRR. \tag{2.5}$$

FAR and FMR are often used interchangeably in the literature, as well as FNMR and FRR [PCK*12]. As in Equ. 2.3 and 2.4, the only difference is that FAR and FRR consider samples failed to be acquired. This work follows the trend in literature and uses FAR and FRR interchangeably with FMR and FNMR as the FTA is considered to equal zero with the work focusing on developing more accurate algorithms rather than systems.

The so-far discussed verification metrics are dependent on the comparison score threshold used to make the genuine/imposter decision. This threshold builds a tradeoff between the FAR and FRR (FMR and FNMR). Therefore, the performance of biometric algorithms is only comparable by fixing one of both error rates and comparing the other. To avoid this and provide one generalized metric, the equal error rate (EER) is used. EER is the common value of FAR or FRR at the operational point (threshold) where they are equal.

This tradeoff between error rates are interesting because different applications require different levels of security and convenience from the same biometric algorithm/system. To present a wide range of performance operation points, receiver operating characteristic (ROC) and detection error tradeoff (DET) curves plots the performance at different decision threshold points.

An ROC curve plots FAR (x-axis) versus TAR (y-axis) (or FMR vs. 1-FNMR) at all possible decision thresholds. This allows the system integrator to informatively choose the threshold that best fits the required security-convenience tradeoff. A DET curve provides the same information as an ROC curve by plotting FAR (x-axis) versus FRR (y-axis) (or FMR vs. FNMR).

2.3.2 Identification performance metrics

Biometric identification systems are categorized into open-set identification and close-set identification. Close-set identification search for the subject identity in a list of identities that is known to contain this subject. Open-set identity systems search for a subject identity in a database that may or may not contain the true identity of the subject.

The cumulative match characteristic (CMC) curve is used in the literature to present the close-set identification performance. CMC plots the number of considered top ranks (x-axis) versus the ratio of identification processes where the correct identity was found in the considered top ranks (x-axis). This ratio will be referred to as the true identification rate (TIR) at a certain rank r.

Open-set identification implies verifying if one of the identities in the reference database belongs to the probed subject. This leads to building a performance metric based on the verification performance of the biometric system and the size of the reference database. Two metrics with a verification-like tradeoff are usually used, the false positive identification rate (FPIR) and the false negative identification rate (FNIR).

While considering the FTA, FPIR for a reference database containing N records is given by

$$FPIR = (1 - FTA) * (1 - (1 - FMR)^N), \tag{2.6}$$

and for a zero FTA and a typically small FMR [LJ11], FPIR can be approximated to

$$FPIR = N * FMR. \tag{2.7}$$

Similarly, FNIR can be approximated to

$$FNIR = N * FNMR. \tag{2.8}$$

Therefore, the errors of an open-set identification system grow linearly with the reference database size. Along with the demand for large scale biometric systems, the motivation for achieving lower verification error rates is higher than ever. This drives the utilization of multi-biometrics to enhance performance, which is a key step in enabling large-scale biometric systems.

2.4 Summary

This chapter discussed the general properties of a typical biometric system. It also presented a high-level view on multi-biometric technologies supported by a look at the related works in the literature. The main metrics commonly used to describe the performance of biometric systems were also discussed in this chapter. Next chapters (3, 4, 5, and 6) will analyze, in more details, this work responses to the previously stated research questions.

3 Pre-fusion processes

Previous chapters motivated and structured the research questions in this work. They also provided basic background knowledge about multi-biometric technologies. This chapter focuses on operations performed in preparation for the information fusion process by tackling two issues. First, this chapter presents a novel normalization technique that tries to align the inference of different score sources by aligning a certain performance-related point in the score distributions. Secondly, this chapter discusses missing data imputation within score-level multi-biometric fusion. This chapter is based on the publications [DON13] and [DFK13].

3.1 Introduction

In order to achieve the goals of multi-biometric fusion, the fusion process expects to receive complete and comprehensible information. The comprehension of different biometric sources should be relatable, and therefore a well-designed normalization of the comparison scores produced by these sources is necessary. On the other hand, and in real operation, some of the fused information might be missing due to problems like a failure to capture. A successful fusion process has to be flexible enough to deal with such cases. In this section, an introduction is built to the later presented work that deals with these two pre-fusion issues.

Starting with normalization, this process adjusts values or measures produced by different sources to a common scale. In multi-biometric systems based on score-level fusion, the values to be normalized are the comparison scores that describe the similarity between a captured biometric characteristic and a stored reference. Similarity scores between captured biometrics and a certain identity reference can be a result of different types of comparisons. These comparisons can be based on different biometric characteristics, algorithms, captures, sensors, or instances of the same characteristic. To build a unified recognition decision from mulit-biometric sources, a fusion process is performed. The fusion can be applied on different levels of the biometric recognition workflow. It can be applied on data, feature, score, or rank-level [DOS13]. Here, score-level fusion is considered as it provides more flexibility to use a wide range of biometric characteristics, sensors, and algorithms (with respect to data and feature-level fusion). Score-level fusion also provides more in-depth information when compared to rank-level fusion.

Comparison scores should be shifted and scaled to become comparable and suitable for fusion, this is performed by score normalization. Score normalization techniques largely affect the performance of multi-biometric systems [JNR05] and thus are an active research field within the study of multi-biometrics. However, conventional normalization techniques only consider the values range with some extended to take into account the nature of the values distributions. No link has been previously made between the performance induced by the scores and their actual values.

This chapter presents a family of novel normalization techniques for score-level multi-biometric fusion. The proposed normalization is not only concerned to bring comparison scores to a common range and scale, it also focuses on bringing certain operational performance points into alignment and thus align score inference. The performance anchored normalization (PAN) algorithms discussed here were tested on the extended multi modal verification for teleservices and security applications database (XM2VTS) [PB06] and proved to outperform

conventional score normalization techniques in most tests. The tests were performed with combination fusion rules and presented as biometric verification performance measures.

The second issue dealt with in this chapter is the imputation of missing data. In the practical use of multi-biometric solutions, biometric sources involved in producing the verification or identification decision do occasionally fail to produce results. This can occur due to the non-universality of biometric characteristics or the low quality of captures especially in ubiquitous systems, such as surveillance and smart living applications.

Missing data can also occur if a subset of the data was not collected in the enrollment process. This particularly happens in large-scale multi-biometric identity management systems as these systems may contain large non-coherent multi-biometric reference databases.

Previous works considered the problem of missing data in multi-biometric systems under the verification scenario. However, the different nature of the identification scenario (1:N comparison), where more than one comparison set produce the decision, imposes challenges on the missing data imputation solutions. This is mainly because of the variable number of missing scores in each comparison involved in the identification (ranking) decision.

This chapter presents the use of support vector regression (SVR) to impute missing score values and compares it to simpler approaches. The effect of different missing data imputation approaches are discussed under both verification and identification scenarios using the Biosecure DS2 score database for development and evaluation. The presented results point out the different imputation methods effects on the verification and identification scenarios. The results also discuss the validity of using complex imputation method.

In the next section 3.2, a review of the related works in the literature is presented. Sections 3.3 and 3.4 present a detailed description of the methodologies and experiments carried out to answer the research questions regarding the two pre-fusion processes, score normalization and missing data imputation. A final discussion concludes this chapter in section 3.5.

3.2 Related work

This section presents an overview on the recent related work dealing with score normalization and missing data imputation in multi-biometric fusion.

As established earlier in the introduction, *score normalization* is an essential step towards successful multi-biometric score fusion. Jain et al. presented a comparison between different score normalization methods for multi-biometric fusion [JNR05]. Other works deeply discussed the effect of normalization on the biometric verification performance with a focus on methods such as zero-mean normalization (z-score), min-max, median absolute deviation (MAD), TanH and decimal scaling normalization [SUM*05, SG07]. Some of these normalization methods will be discussed in more detail in Section 3.3.1.

Min-max normalization is a simple technique to rescale the range of data to fit into a [0,1] range and it only depends on the minimum and maximum values of the training data, however, this makes it vulnerable to outliers. Min-max normalization does not consider the nature of the distribution of the values nor the actual inference of these values on the application in hand. Due to its simplicity, min-max normalization was used regularly in score-level multi-biometric fusion [JNR05]. The MAD normalization tries to capture information about the values distributions by considering their median and median absolute deviation assuming a Gaussian distribution of the data. MAD was popularized by the work of Hampel in 1974 [Ham74, LLK*13]. Just like MAD normalization, z-score normalization tries to unify the normalized values distribution based on their standard deviation and mean value [Kre00]. Z-score transforms the values to have a standard deviation of one and a mean value of zero, however, no relation to these values inference is considered.

TanH was introduced by Hample et al. [HRRS86] and is based on the standard deviation and mean value of the genuine comparisons values in a way that reduces the influence of the points at the tails of the distribution. Just like the previously presented normalization approaches, TanH constructs no clear link between the normalized values and their influence on the performance and thus their real effect in a fused system. These approaches are widely used to normalize comparison scores from different biometric sources as an important step in score-level multi-biometric fusion [SUM*05, SG07, JNR05]. However, these previous works did not build any direct link between the values after normalization and their inference on the performance, and thus their real influence on the final biometric decision.

Regarding *missing data imputation* in multi-biometric fusion, previous works dealt with the problem, although, most of the works considered only the verification scenario. Fatukasi et al. [FKP08] proposed a k-nearest neighbor approach to estimate the missing score. This approach was compared to mean and zero imputations along with the regression approach used by Little in [Lit92]. The results showed that the k-nearest neighbor approach and the simple mean imputation performed best and was dependent on the degree of correlation between the fused biometric sources.

Poh et al. [PWM*10] suggested an approach based on support vector machines (SVM) with neutral point substitution. Different versions of the solution were tested and compared to raw fusion under the verification scenario. Equal error rate (EER) values were reported as performance measure for evaluation on a subset of the Biosecure database [PBK10], which proved the validity of the proposed solution. This approach replaced the missing data with a neutral point value that will have no effect on the SVM decision, limiting the scope of the fusion to using SVMs.

In the work of Ding [DR10], the use of maximum likelihood estimation was evaluated along with two variations of the Gaussian mixture model (GMM) and the multiple imputation proposed by Rubin [Rub87]. Results were reported as verification performance achieved on the Michigan state university dataset (MSU). The maximum likelihood estimation performed poorly in comparison to GMM based approaches. However, unexpectedly, training GMM approaches using smaller amount of training data achieved significantly better performances. More recently, Ding and Ross [DR12] presented a comparison between imputation methods for missing scores. The comparison included approaches just as the k-nearest neighbor estimation, maximum likelihood estimation, GMM, predictive mean matching, and multiple imputation using the Markov chain Monte Carlo and the chained equation. The different approaches were evaluated on the MSU database [RJ03] as verification performance with superiority to the GMM based approaches.

All previously mentioned works considered the problem of missing data in multi-biometric systems under the verification scenario. The nature of the identification tasks is fundamentally different from the verification tasks. Identification considers a number of comparisons to produce a final decision (or recommendation), unlike the verification decision based on one comparison. This makes the identification scenario more challenging to missing data imputation as it faces a variable number of missing scores in each task.

Nandakumar et al. [NJR09] explicitly dealt with the missing data under the identification scenario. However, the likelihood ratio-based score fusion approach presented was designed to be robust to missing data instead of focusing on missing data imputation.

The next sections of this chapter will go deeper into the presented solutions and their evaluation.

3.3 Performance anchored score normalization

This section presents a family of novel normalization techniques for score-level multi-biometric fusion. The proposed normalization approaches focus on bringing certain operational performance points into alignment and thus align the score inference, rather than only bringing comparison scores to a common range and scale.

3.3.1 Baseline normalization

As mentioned before, comparison scores processed by the fusion algorithm are usually not homogeneous as they are produced by different sources. Therefore, these scores have to be normalized before fusing. Some of the most common normalization techniques are min-max, z-score , TanH, and MAD normalization [JNR05].

The parameters that define the normalization process are usually determined based on the statistical properties of the biometric system in hand. The performances of normalization techniques are not directly comparable as they depend on the overall multi-biometric system. Therefore, the normalization performance is measured based on the performance of the multi-biometric system. In the following, more details are presented on the baseline normalization techniques used in the later discussed experiment (Section 3.3.3).

Given the development comparison scores set S_k, $k = 1,2,\ldots,N$, for each biometric source, the normalized score S' is defined as a function of the score S. The min-max normalization does not depend on the distribution properties and it considers only the range of the scores. This aims at mapping the raw scores into the range of [0,1]. The normalized score by min-max normalization is given by

$$S'_{min-max} = \frac{S - min\{S_k\}}{max\{S_k\} - min\{S_k\}}. \tag{3.1}$$

The min-max normalization scheme is highly sensitive to outliers [JNR05] as it depends on single minimum and maximum values.

Z-score normalization uses the arithmetic mean (μ) and the standard deviation (σ) of the development score set as parameters. This method assumes a Gaussian distribution of the score values. The z-score normalization showed good performance in many studies [NS09, VIM*07]. The normalized score by z-score normalization is given by

$$S'_{z-score} = \frac{S - \mu}{\sigma}. \tag{3.2}$$

The MAD normalization method uses the median and median absolute deviation instead of the mean and standard deviation used in z-score normalization. This method also assumes a near Gaussian distribution of the comparison score values but is more robust to outliers. The median absolute deviation normalization is given by

$$S'_{MAD} = \frac{(S - median)}{MAD}, \tag{3.3}$$

$$MAD = median(|\{S_k\} - median|). \tag{3.4}$$

The last normalization baseline algorithm used here is the TanH normalization [JNR05]. The TanH normalization is formulated as

$$S'_{TanH} = 0.5\{\tanh(0.01(\frac{S-\mu_G}{\sigma_G})) + 1\}, \tag{3.5}$$

where μ_G and σ_G are the mean and standard deviation of the genuine scores distribution (of training data).

The selection of a proper normalization method is a tradeoff between efficiency and robustness. Here, only normalization methods that require no special parameter tuning were considered. Methods like double sigmoid normalization [CMM00] and TanH normalization based on Hampel estimators [HRRS86] require the fine tuning of certain parameters, while in this work the focus is on normalization methods that depend only on parameters that can be simply acquired from the development data statistics.

3.3.2 Proposed score normalization

The proposed normalization techniques do not only consider the range and scale of comparison scores, they also align the scores of different biometric sources with respect to a certain performance operation point. The score value that aligns the score distributions is called the anchor, hence the notion of performance anchored normalization (PAN). This anchor score value is obtained based on the statistics of the development data. The anchor value considered in this work is the score threshold value at the EER operation point, and is noted here by TH_{EER}. Applying this threshold value to separate genuine and imposter scores produces an equal false acceptance rate (FAR) and false rejection rate (FRR).

This operating point alignment aims at the alienation of the undesired weight-like effect embedded in score value distributions, even after conventional normalization (e.g. min-max normalization). For example, if two comparison score sources have score distributions in the same range. However, one of the sources has a distribution that is more shifted to the high value side of the range, the scores of this source will have a higher weight when used with combination fusion rules. This higher weight effect might be an incorrect assumption by the system and leads to a lower performance.

The first PAN technique presented here is the PAN-min-max normalization. Here, the min-max normalization is extended by anchoring the middle point of the score range at the EER operation point TH_{EER}. The PAN-min-max normalized score $f(S)$ is given by

$$S'_{PAN-min-max} = \begin{cases} \frac{S-min\{S_k\}}{2(TH_{EER}\{S_k\}-min\{S_k\})} & if S <= TH_{EER} \\ 0.5 + \frac{S-TH_{EER}\{S_k\}}{max\{S_k\}-TH_{EER}\{S_k\}} & if S > TH_{EER} \end{cases} \tag{3.6}$$

A modified MAD normalization is also presented here. The PAN-MAD normalization considers the anchor value (TH_{EER}) as a pivot value instead of the median of development scores used in the conventional MAD normalization. The PAN-MAD normalization is given by

$$S'_{PAN-MAD} = \frac{(S-TH_{EER})}{MAD_{PAN}}, \tag{3.7}$$

where the MAD_{PAN} is formulated as

$$MAD_{PAN} = median(|\{S_k\} - TH_{EER}|). \tag{3.8}$$

As for the PAN-MAD normalization, a performance anchored version of TanH normalization is presented. The PAN-TanH normalization here considers the TH_{EER} as anchor score value and is given by

$$S'_{PAN-TanH} = 0.5\{\tanh(0.01(\frac{S - TH_{EER}}{\sigma_G})) + 1\}, \tag{3.9}$$

where σ_G is the standard deviation of the genuine scores distribution.

In Figures 3.1 and 3.2, the distribution of normalized genuine and imposter evaluation comparison scores resulted from different normalizers are shown. Comparisons from two sources are visualized, the face matcher based on discrete cosine transform and Gaussian mixture model (DCTb-GMM) and the voice matcher based on the linear filter-bank cepstral coefficient and Gaussian mixture model (LFCC-GMM) matchers of the XM2VTS LP1 database [PB06]. It can be noticed that the PAN normalization techniques did align the score distribution of both biometric sources at the anchor value TH_{EER}, the EER occurs inside the area of overlap between genuine and imposter distributions. This alignment can be noticed in the face-voice distribution pairs normalized by PAN methods (d and j, e and k, f and l) in Figures 3.1 and 3.2.

3.3.3 Experimental setup

For the development and evaluation of the proposed solution, two parts of the XM2VTS multi-biometric database were used, LP1 and LP2 [PB06]. LP1 and LP2 contain comparison scores by different face and voice baseline experts. The score sets are split into evaluation and development sets. LP1 contains eight score sources (5 face experts and 3 voice experts) while LP2 contains five sources (2 face experts and 3 voice experts). The major difference is that LP1 used three training captures per client while LP2 used four. For more details about the XM2VTS score database, one can refer to the work of Poh et al. [PB06].

Score normalization parameters for each biometric source and for both LP1 and LP2 parts of the database were obtained from the development data. Normalization parameters were acquired for the four baseline normalization (z-score, min-max, TanH and MAD) and the three proposed PAN normalization techniques (PAN-min-max, PAN-MAD, PAN-TanH). Normalization was performed on the evaluation data using all the seven techniques producing seven normalized evaluation sets.

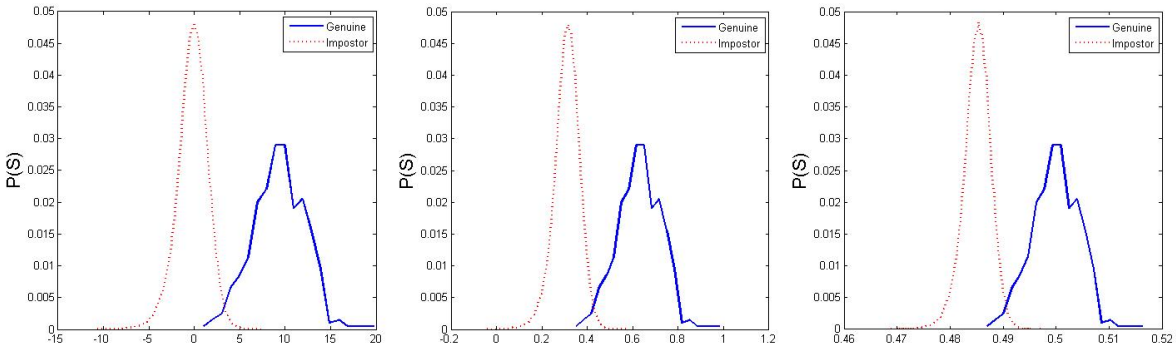
After normalization, combination fusion rules were used to fuse the normalized comparison similarity scores. Three combination fusion approaches were evaluated, the simple sum rule, and the weighted-sum rule with two different weighting approaches. Assuming the fused score F is a function of the N biometric scores, the combination rules can be formulated as follows

$$\text{The sum rule: } F_{sum} = \sum_{k=1}^{N} S_k, \tag{3.10}$$

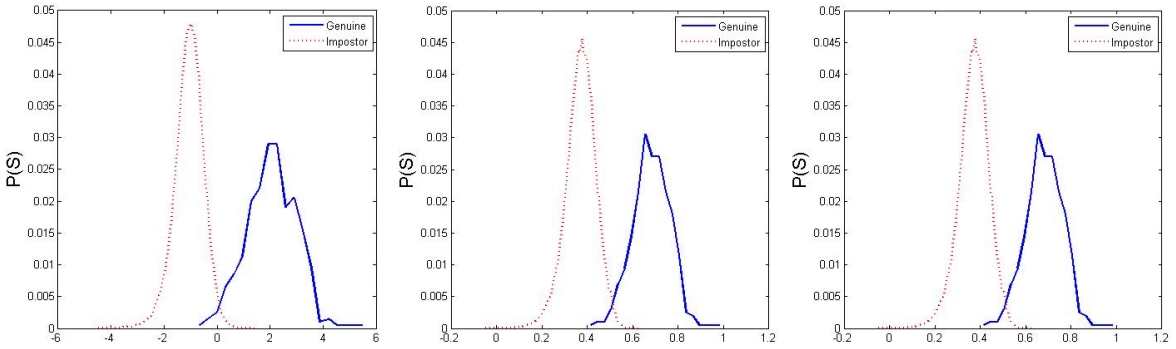
$$\text{The weighted-sum rule: } F_{weighted-sum} = \sum_{k=1}^{N} w_k S_k, k = \{1, \ldots, N\}, \tag{3.11}$$

where the weight of each source w_k is given by equal error weighting (EERW) and overlap deviation weighting (OLDW), both discussed in details in Chapter 4 and calculated in the Equations 4.1 and 4.11

Many works discussed the performance of different basic combination rules, with the sum rule usually producing higher performance [CBFC04]. However, the performance of combination rules depends on the normalization used. Different pairing between combination rules and normalization techniques produces varying results [JNR05].

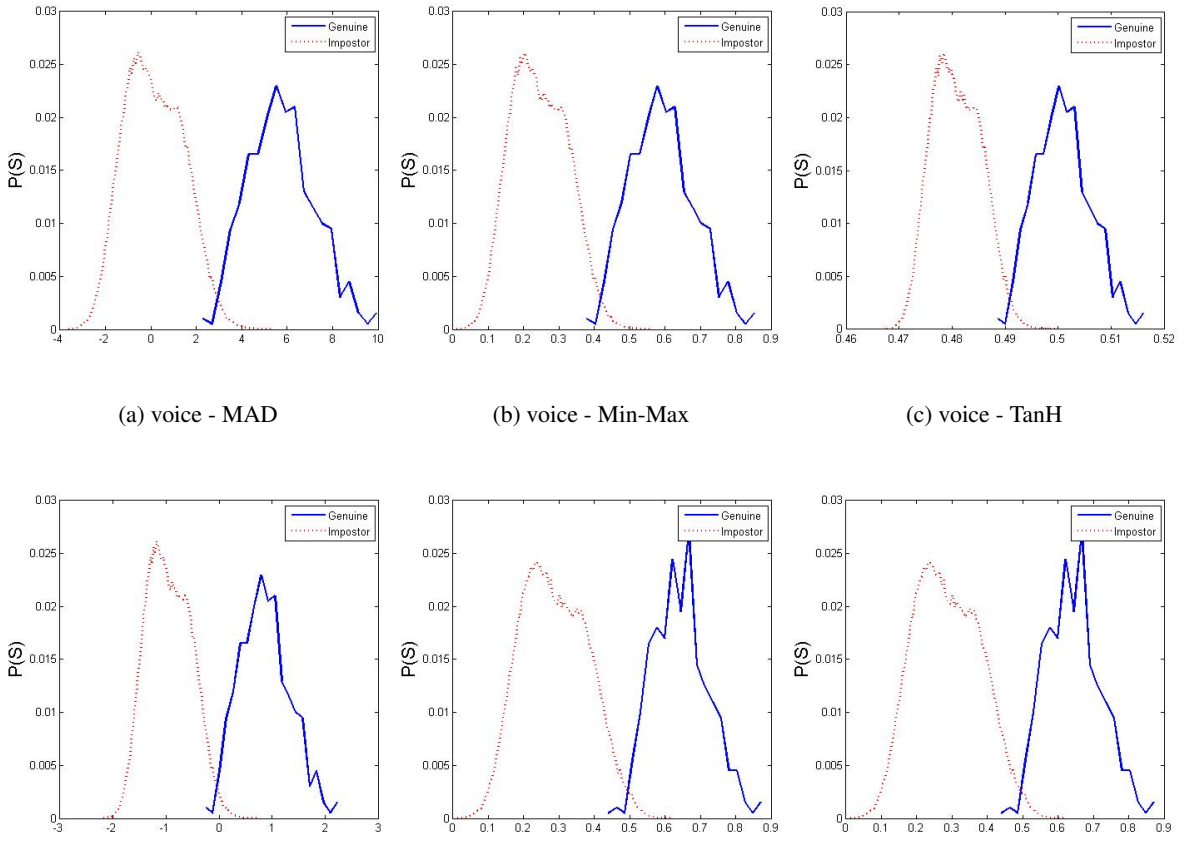


(a) face - MAD (b) face - Min-Max (c) face - TanH



(d) face - PAN-MAD (e) face - PAN-Min-Max (f) face - PAN-TanH

Figure 3.1: genuine (blue) and imposter (red) score distribution of normalized scores for face (DCTb-GMM) [PB06] using different normalizers.



(a) voice - MAD (b) voice - Min-Max (c) voice - TanH

(d) voice - PAN-MAD (e) voice - PAN-Min-Max (f) voice - PAN-TanH

Figure 3.2: genuine (blue) and imposter (red) score distribution of normalized scores for voice (LFCC-GMM) [PB06] using different normalizers.

		sum		EERW-sum		OLDW-sum	
		0.01%FAR	0.001%FAR	0.01%FAR	0.001%FAR	0.01%FAR	0.001%FAR
XM2VTS-LP1	min-max	3.50%	8.00%	3.50%	4.75%	**1.00%**	**2.25%**
	MAD	11.25%	13.50%	12.50%	14.75%	**1.00%**	**2.25%**
	TanH	4.00%	8.50%	2.50%	4.50%	4.00%	6.75%
	z-score	**2.50%**	**5.75%**	**1.25%**	**3.25%**	**1.00%**	**2.25%**
	PAN-min-max	3.50%	6.25%	1.50%	4.25%	**1.00%**	**2.00%**
	PAN-TanH	**1.00%**	**3.75%**	**0.75%**	**2.00%**	**1.00%**	**2.25%**
	PAN-MAD	3.00%	6.25%	3.75%	4.50%	**1.00%**	**2.25%**
XM2VTS-LP2	min-max	2.75%	3.75%	3.75%	4.25%	**0.25%**	**0.25%**
	MAD	12.00%	19.00%	12.00%	17.50%	**0.25%**	**0.25%**
	TanH	2.25%	**2.50%**	2.50%	**3.00%**	1.00%	2.25%
	z-score	**2.00%**	3.50%	2.75%	3.75%	**0.25%**	**0.25%**
	PAN-min-max	**1.25%**	**2.75%**	**2.00%**	**3.50%**	**0.25%**	**0.25%**
	PAN-TanH	**1.25%**	3.25%	**1.50%**	**3.50%**	**0.25%**	**0.25%**
	PAN-MAD	2.75%	3.75%	3.25%	4.25%	**0.25%**	**0.25%**

Table 3.1: FRR values achieved at fixed FAR for the different experiment settings on two databases. The best two rates across normalization techniques (columns) are in bold. Notice the consistent improvement in performance when the normalization approaches are anchored by a performance related point (PAN).

3.3.4 Results

The performance of the normalization techniques were compared by considering the multi-biometric fusion results under verification scenario. The performance was reported as FRR at fixed FAR for each normalization and fusion rule pairing. This allows to discuss the performance at different operational points that might be of interest for different applications/users. The achieved error rates are presented in Table 3.1

From Table 3.1, as expected, the weighted-sum fusion approaches scored lower error rates as they induce more control over the relative influence of different sources in the multi-biometric decision. All the PAN approaches achieved significantly lower error rates than their originally, not anchored, versions in most experimental settings. Under the sum rule fusion, min-max normalization achieved 21% and 26% lower FRR rates at 0.001%FAR when anchored (PAN-min-max), on the LP1 and LP2 databases. At the same FAR value and fusion rule, the performance anchored PAN-MAD achieved 53% and 80% lower FRR rates compared to the MAD normalization.

Fusing using the OLDW-sum rule achieved comparable results between different normalization approaches as they reach performance saturation induced by the weighting approach. The widely used z-score normalization achieved good performances. However, it was outperformed by the PAN methods on all experimental settings. The results proved the benefit of the proposed normalization approaches by consistently enhancing the multi-biometric verification performance.

3.4 Missing data imputation

This section discusses solutions for missing data imputation in multi-biometric score-level fusion. A missing data imputation solution based on support vector regression is presented in this work and compared to four baseline solutions. The evaluation is then carried under both the verification and the identification scenarios in an effort to show the effect of missing data imputation on the relatively understudied multi-biometric identification scenario.

3.4.1 Methodology

In this section, the proposed SVR missing data imputation approach is presented along with four baseline approaches dealing with the missing data in multi-biometric systems.

Weighted-sum score-level fusion was used to produce the multi-biometric fused score. Under the weighted-sum algorithm, each biometric source is assigned a weight that aims at minimizing the verification error. The fused score F is given as the sum of the product of weights and their corresponding scores as described in Equation 3.11.

Many cases occur where one or more of the summed scores in a comparison is missed, and thus the previously mentioned fusion rule produces undependable results.

In an effort to impute missing values in a given multi-biometric comparison, different approaches were implemented and evaluated. In the following, the *min*, *max*, *median*, and *SVR* approaches are formulated along with, the more conventional, weight distribution approach.

Weight distribution: given a multi-biometric system that is based on N biometric sources S_k, with each having the weight w_k, $k = \{1,\ldots,N\}$. Using weight distribution, a new weight distribution will be assigned when one or more score values are missed from a comparison. These new weights are noted by w'_k. In the following E and M are subsets of k where scores *exist* or *missed* subsequently.

Weight distribution is used to deal with missing scores by distributing the weights of the missing scores on the existing scores. This distribution is performed proportionally to the initial weights of the existing scores. The new weights after weight distribution is given as

$$w'_k = \begin{cases} w_k + \frac{w_k \sum_{m\in M} w_m}{\sum_{e\in E} w_e} & if\, k \in E \\ Zero & if\, k \in M \end{cases} \tag{3.12}$$

Imputation by minimum rule: in a more conservative approach, the minimum existing score value in a multi-biometric comparison scores set is assigned to the missing value(s) S_m in it. Here, S_m and S_e denote the missing and existing scores in a comparison subsequently, $m \in M$and $e \in E$. The imputed missing score S'_m can be formulated as

$$S_{I-min} = min_{e\in E} S_e. \tag{3.13}$$

Imputation by maximum rule: another approach assigns the maximum existing score in the comparison to the missing values. The imputed missing scores are given by

$$S_{I-max} = max_{e\in E} S_e. \tag{3.14}$$

Imputation by mean rule: the *mean* approach calculates the arithmetic mean value of the existing scores in a comparison and assigns this value to the missing values.

$$S_{I-mean} = mean_{e \in E} S_e. \tag{3.15}$$

Imputation by SVR: in this work, the use of SVR for missing score imputation is proposed. SVR [DBK*96] is an extension to the SVM that performs regression estimations. SVR is implemented here to impute missing comparison score values based on the existing ones.

SVR estimates a continuous value function that maps a set of input values into an output value. This function is optimized through a training process. From this point of view, SVR performs similarly to neural networks. However, SVR performs structural risk minimization in the regression and therefore achieves global optimization where neural networks achieve local optimization [Eub99].

A different SVR was trained to impute each biometric source scores under all possible missing data conditions. These conditions range from the case where all other scores being available to the case where only one other source score value is available (different case for different available sources). For example, an SVR was trained to impute the first biometric source score when all other sources are available. Another SVR was trained to impute the first biometric source when only the third and fourth sources are available, etc.

SVR, just like SVM, tries to minimize the generalization classification error. This is done by calculating optimal regression coefficient to fit the so called $\xi - intensive$ band. This is optimized so that a penalty is imposed if a training sample does not fit this band. This penalty corresponds to the distance between the $\xi - intensive$ band and the sample.

In the next section, the carried experiments are discussed along with the used database. The results achieved by the presented missing score imputation approaches are discussed under both, verification and identification scenarios.

3.4.2 Experimental setup

This section presents the used database and the experiment procedure for both verification and identification scenarios. The results achieved by different approaches are also presented along with a discussion.

Data: the Biosecure DS2 score database [PBK10] was used to develop and evaluate the proposed solution. Here, a subset of the data is used containing three fingerprint matchers and one face matcher. The database contains a total of 207 clients split into 51 clients for training and 156 clients for evaluation. It must be mentioned that cross-sensor comparison scores are included in the database.

The face comparison considered was based on frontal face images captured without flash and is noted in the database by Fnf1. While the three fingerprint comparisons were of fingers from the left hand (thump, index and middle finger). The fingers were optically scanned by direct contact. The fingerprint channels used here noted in the database by Fo1, Fo2, and Fo3.

Missing data simulation: missing values were introduced into the database by randomly removing scores until the desired percentage of missing data is reached.

That was performed by constructing the database as an Nx4 matrix. Each raw represents one comparison set between two identities. Each column represents a biometric source. Scores were removed randomly from this matrix until the percentage of missing data is reached. The only constrains was that each comparison set (row) must contain at least one existing score left.

Normalization: min-max normalization was performed to bring comparison scores produced by different biometric sources to a comparable range. Min-max normalized score is given in Equation 3.1.

Weighting: different biometric sources have different performance and thus must have varying role in the fusion process, this is influenced by assigning a weight to each biometric source. In this work, the gradient descent approach presented by Basak et al. [BKTR10] is used to calculate optimized weights for different biometric sources. This approach optimizes the weights such that the weighted-sum of genuine scores is higher than the weighted-sum of the imposter scores.

3.4.3 Results

Experiments were performed under two scenarios, verification and identification. The effects of different missing score imputation approaches are presented as the performance of the verification and identification multi-biometric systems.

Verification: under verification, the scores were normalized and each comparison set was fused using the previously mentioned weighted-sum rule. This was done on two sets of data, one with 20% missing data and the other with 40% missing data. The results were presented as receiver operating characteristic curves (ROC) by changing the decision threshold from minimum to maximum and calculating the true acceptance rate (TAR) and the FRR at each decision threshold. Verification performance results for different missing data imputation approaches are shown in Figures 3.3 and 3.4 for the cases of 20% and 40% missing data.

From Figures 3.3 and 3.4, one can notice the high verification performance when the *mean* and *SVR* missing data imputation approaches are used. The conventional *weight distribution* approach falls behind in performance, however, it outperforms the *minimum* and *maximum* missing data imputation approaches. It can be noticed that more sophisticated solution, such as the one based on *SVR*, does not provide a significant advantage over the best performing simpler solution (*mean*).

Identification: knowing that the biometric verification is defined as the use of biometric characteristics to confirm or reject a claimed identity of a person (1:1 comparison), the biometric identification differs as it depends on a larger number of comparisons (1:N comparison) in an effort to assign an identity to an unknown captured individual.

Multi-biometric identification systems (1:N comparison) impose many challenges on the fusion process in practice when compared to the verification scenario. These challenges range from big databases to varying references and captures qualities. Missing scores largely affect the fusion process under the identification scenario as the decision depends on a large number of comparison sets (one comparison set in verification). These different comparison sets, affecting the identification decision, have different missing data characteristics. However, the fusion output of each of these comparisons must be comparable in order to produce a well ranked identification decision.

In the carried experiment, the procedure in the work of Basak et al. [BKTR10] was followed. Initially, what is called a "superset" of comparisons was created. This superset contains all identities appearing in the top k matches (highest scores) of each biometric source. For example, if an identity appears in the top k matches of the face matcher, all the comparison (including all biometric sources scores) belonging to this identity is added to the superset, even if this identity does not appear in the top k matches of any other biometric source. The k value used in this work is ten ($k = 10$).

All comparisons in the superset were fused using the weighted-sum fusion rule. This resulted in a list of fused scores that was later ranked. The identification performance was visualized as cumulative match characteristic (CMC) curves. In a CMC curve, the horizontal axes represent the top ranks considered. The vertical axes

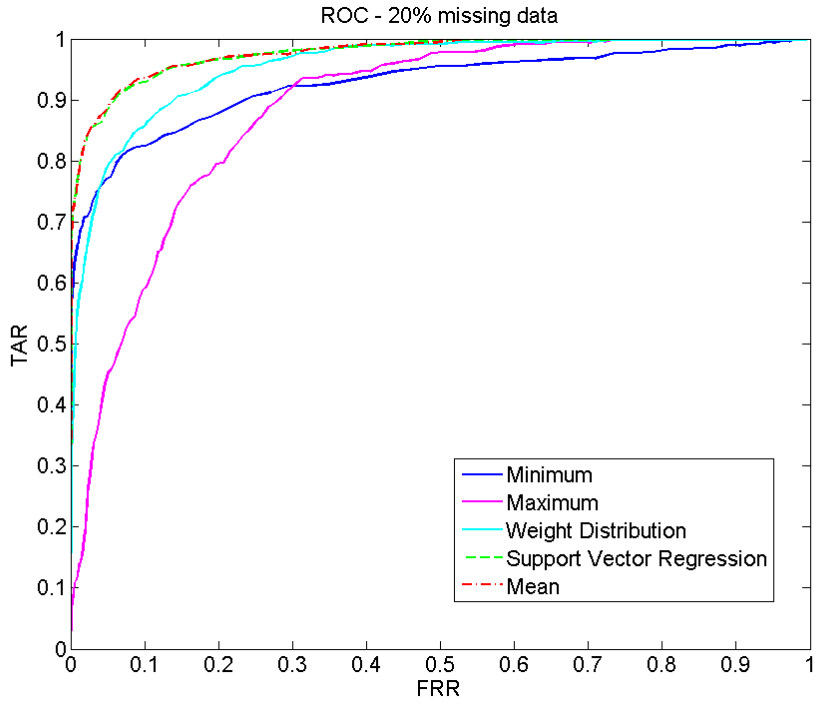


Figure 3.3: missing data imputation under verification scenario: ROC curves produced by different missing data imputation approaches given 20% missing data.

represent the rate of tests where the correct identity lied in the considered top ranks (horizontal axes). The identification performance with different missing value imputation approaches is shown in Figures 3.5 and 3.6 for the cases where 20% and 40% of the data is missed.

In Figures 3.5 and 3.6, the performance superiority of the *mean* and *SVR* missing data imputation approaches is clear. Where the *minimum* approach follows closely especially with lower percentage of missing data. Given less information (more missing data), an extreme imputation approach, such as the minimum approach, is expected to be less generalized and sensitive to outliers. The weight distribution and the *maximum* approaches result in lower performance especially for the important high ranks.

It can be noticed that the *minimum* approach performs well under the identification but not under the verification scenario. Moreover, the weight distribution approach provides acceptable results under verification but performs poorly under the identification scenario. This is due to the different operational nature of both scenarios.

In verification, a fused score resulting from a single comparison produces the decision based on a given imposter/genuine threshold. On the other hand, identification decision depends on a set of fused scores produced from different comparison (with different missing data characteristics). This set of fused scores is compared and ranked to produce the final identification result.

For example, a fused score of a certain comparison (between captured and reference characteristics of the same individual) is considered correct under closed-set identification if it is the highest fused scores within all

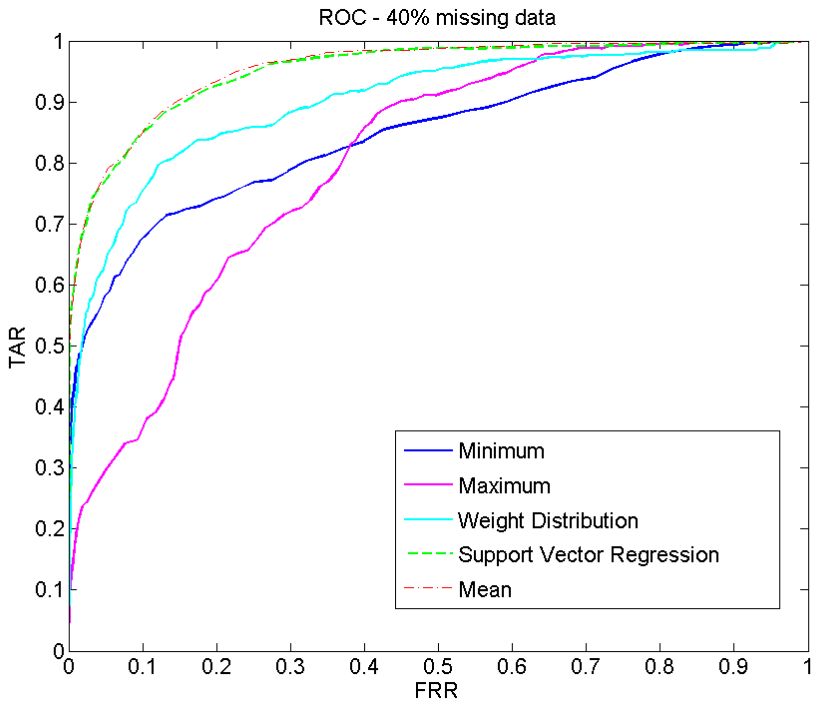


Figure 3.4: missing data imputation under verification scenario: ROC curves produced by different missing data imputation approaches given 40% missing data.

the comparisons between the captured individual and all references. This is considered regardless of the value of the fused score (only the rank). However, under verification, the biometric decision is correct (genuine in this case) only if the fused score is over a certain similarity threshold, regardless of other comparisons scores. Therefore, different approaches handeling with missing data perform relatively different under both scenarios.

3.5 Summary

This chapter discussed operations performed to prepare biometric information to be fused within the multi-biometric decision-making process. Two aspects were in focus, biometric comparison score normalization, and missing data imputation.

At first, this chapter presented novel score normalization techniques for multi-biometric score-level fusion. Previously proposed normalization approaches focused on transferring score values from different sources into a common range. More advanced approaches also considered the unification of the distribution of these values. However, no work has been done to build a normalization approach that considers the performance induced by these scores, and thus their interpretation. The proposed normalization approach in this chapter successfully introduced a modification into three of the standard normalization approaches to introduce an alignment of a

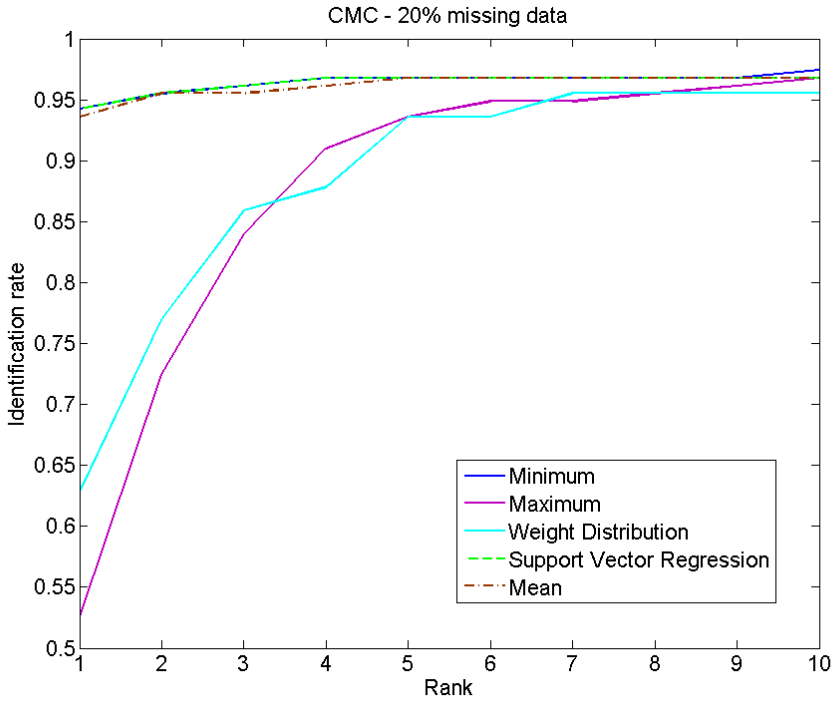


Figure 3.5: missing data imputation under identification scenario: CMC curves produced by different missing data imputation approaches given 20% missing data.

performance related score value, the anchor. This achieved performance anchored normalization was visually presented in genuine/imposter distributions of different sources. A quantitative proof of the validity of the proposed approach was also presented by evaluating the multi-biometric verification performance on the XM2VTS multi-biometric database. This evaluation pointed out the enhanced performance induced by the proposed normalization when compared to the baseline solutions, e.g. the FRR values were reduced by 21% to 26% when the performance anchor was introduced to the min-max normalization under the sum rule fusion for the XM2VTS-LP1 and XM2VTS-LP2 databases.

The second aspect of this chapter is the imputation of missing data in score-level multi-biometric fusion. Due to failures to capture or the non-universality of biometric characteristics, multi-biometric probes or references might fail to produce a capture of enough quality. This results in multi-biometric comparisons where some sources do not provide comparison scores. This problem has been studied earlier under the verification scenario where the decision is based on one comparison. However, the effect on identification operations including a large number of comparisons and different missing data scenarios in each of these comparisons was still understudied. This chapter evaluated the effect of different imputation solutions under both, the verification and identification, scenarios. This uncovered the different effects of these imputation approaches in these operations modes, e.g. the minimum imputation rule has a very negative effect on the verification performance while it achieves some of the best identification results. A more sophisticated approach based on support vector regression was also

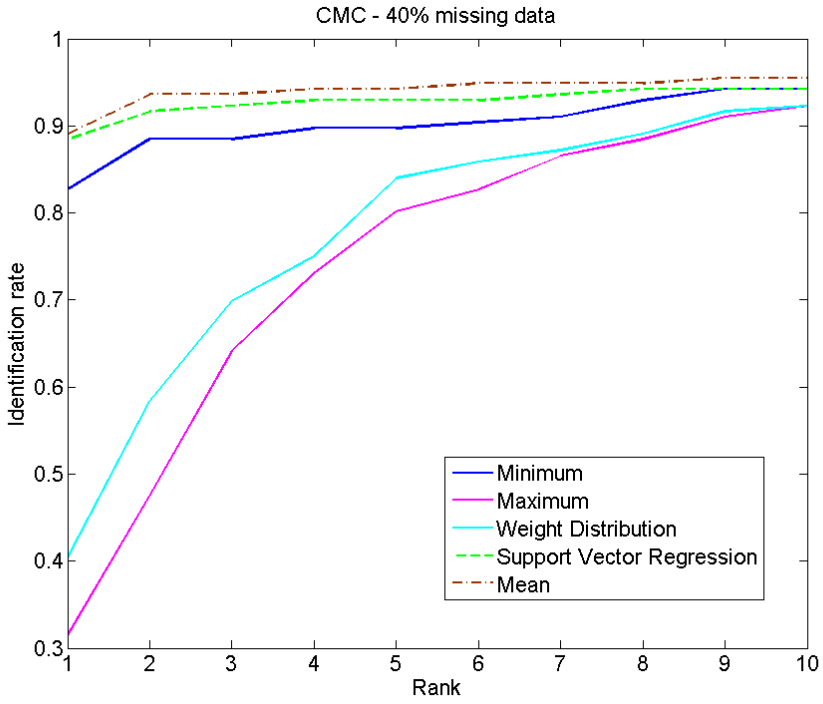


Figure 3.6: missing data imputation under identification scenario: CMC curves produced by different missing data imputation approaches given 40% missing data.

proposed for missing data imputation. Although this approach performed well under both operation scenarios, it did not improve the performance achieved by simpler solutions (mean rule) significantly.

This chapter presented an answer to the *RQ1* by designing a performance related score normalization approach and proving its positive effect on the multi-biometric decision accuracy. It also responded to the *RQ2* by analyzing the effect of different missing data imputation approaches on the multi-biometric verification and identification performances and proposing a regression based approach to explore its influence on the multi-biometric performance. Next chapter will focus on a different aspect in this work, the fusion process optimization. In particular, the weighting of multi-biometric sources within score-level fusion.

4 Multi-biometric source weighting

Chapter 3 was concerned with pre-fusion processes, namely score normalization and missing data imputation. This chapter focuses on optimizing the fusion process by discussing the score-level multi-biometric source weighting. This is done by presenting different weighting approaches that controls the relative effect of each biometric source in the fusion process. Two paths are taken to achieve this goal. The first presents a weighting technique that considers the properties of the genuine and imposter scores distributions to capture the overall performance of the biometric source, as well as an indication of its confidence. The second approach focuses on deriving the biometric source weights from an identification performance representation, the cumulative match characteristic curve. This chapter is based on the published papers [DON14a] and [DON14b].

4.1 Introduction

Multi-biometrics tries to use multiple biometric information sources to enhance performance and to overcome the limitations of the conventional uni-modal biometrics. Such limitations are noisy data, low distinctiveness, intra-user variation, non-universality of biometric characteristics, and vulnerability to spoof attacks. Information fusion is used to produce a unified biometric decision based on multiple biometric sources. Simple approaches such as the sum rule score-level fusion proved to achieve high performance compared to more sophisticated approaches [RJ03]. A step ahead is the weighted-sum rule where each biometric source is weighted to indicate its relative importance, and thus contribution, to the final fused biometric decision.

Searching for the optimal weights combination can be done by exhaustive search to find optimal solution on training data. However, as will be shown in the next sections, this solution has low generalization ability. This chapter presents two different paths to assign optimal weights to different biometric sources within a score-level multi-biometric system. These weights are utilized in the effective and widely used weighted-sum fusion rule to produce multi-biometric decisions.

The first discussed solution is based on the characteristic of the overlap region between the genuine and imposter scores distributions to capture the degree of confidence in the biometric source decisions. It also integrates the performance of the biometric source represented by its equal error rate. This solution aims at avoiding the shortcomings of previously proposed solutions such as low generalization abilities and sensitiveness to outliers. The proposed solution is evaluated along with the state-of-the-art and best-practice techniques. The evaluation was performed on two databases, the biometric scores set BSSR1 (BSSR1) and the extended multi modal verification for teleservices and security applications data7base (XM2VTS) and a satisfying and stable performance was achieved.

The second proposed multi-biometric weighting approach tries to investigate the properties of the cumulative match characteristic (CMC) curve, which represents the biometric performance under the identification scenario, and extract biometric source weights based on these properties.

The CMC curve represents the performance of a biometric system within a close-set identification scenario. This work investigates the properties of the CMC curve and how it can be utilized to assign optimized weights to different biometric sources within a weighted-sum score-level multi-biometric fusion solution. A number of

CMC curve properties are defined in this work. These properties are later used to calculate relative weights for the fused multi-biometric sources. The proposed solution is evaluated along with a set of state-of-the-art and best-practice weighting techniques. The evaluation was performed on the BSSR1 database and a satisfying and stable performance was achieved.

In the next section 4.2, a brief review of the related works in the literature is presented along with the considered baseline weighting approaches. Sections 4.3 and 4.4 present a detailed description of the two proposed paths into multi-biometric weighting with experimental comparisons to the baseline solutions. A final discussion concluded this chapter in section 4.5.

4.2 Related work

Weighting biometric sources controls their relative effect in the fused decisions. These weights are often related to the performance of these individual biometric sources. Weighting methods based on the statistics of the imposter and genuine scores distributions showed better and more generalized performance when compared to weights optimized by exhaustive search [VT12]. Weighting based on the equal error rate (EER) of biometric sources is widely used [JNR05] along with approaches based on D-Prime calculations [SUM*05] and Fisher discriminant ratio (FDR) [PB04].

A comparative study by Chia et al. [CSN10] discussed the performance of the most common weighting approaches and proposed a weighting algorithm that depends on the width of the overlapped area between the imposter and genuine scores distributions. Other works proposed fusion approaches based on non-linear combiners [PB03]. Benchmarking quality-based multi-biometric fusion was also discussed by Poh et al. [PBK*09].

Other Studies considered the properties of the receiver operating characteristics (ROC) curve to optimize the multi-biometric fusion process. These works mostly considered the maximization of the area under the ROC curve such as in the works of Toh et al. [TKL08] and Villegas et al. [VP09].

In the following, a list of biometric source weighting approaches is presented. These approaches present the state-of-the-art and common practices in multi-biometric fusion using the weighted-sum rule.

a) EER weighted (EERW): equal error rate is the common value of the false acceptance rate (*FAR*) and the false rejection rate (*FRR*) at the operational point where both *FAR* and *FRR* are equal. EERW was used to linearly combine biometric scores in the work of Jain et al. [JNR05]. The EER is inverse proportional to the performance of the biometric source. Therefore, for a multi-biometric system that combines N biometric source, the EER weight for a biometric source k is given by

$$w_{k\_EERW} = \frac{\frac{1}{EER_k}}{\sum_{k=1}^{N} \frac{1}{EER_k}}. \tag{4.1}$$

b) D-Prime weighted (DPW): D-Prime is used to measure the separation between the genuine and the imposter scores [SUM*05]. High separation indicates a higher performance of the biometric source. Given that σ_k^G and σ_k^I are the genuine scores and imposter scores standard deviations and μ_k^G and μ_k^I are their mean values, the D-prime is given by

$$d'_k = \frac{\mu_k^G - \mu_k^I}{\sqrt{(\sigma_k^G)^2 + (\sigma_k^I)^2}}, \tag{4.2}$$

and it is directly proportional to the performance of the biometric source and thus the weight can be calculated as

$$w_{k\_DPW} = \frac{d'_k}{\sum_{k=1}^{N} d'_k}. \tag{4.3}$$

c) NCW weighted (NCWW): the non-confidence width (NCW) weight was proposed by Chia et al. [CSN10] to weight biometric sources for score-level multi-biometric fusion. NCW corresponds to the width of the overlap area between the genuine and imposter scores distributions. Given that Max_k^I is the maximum imposter score and Min_k^G is the minimum genuine score, NCW is given by

$$NCW_k = Max_k^I - Min_k^G, \tag{4.4}$$

as the NCW is inverse proportional to the biometric source performance, the weights based on the NCW are given as

$$w_{k\_NCWW} = \frac{\frac{1}{NCW_k}}{\sum_{k=1}^{N} \frac{1}{NCW_k}}. \tag{4.5}$$

d) FDR weighted (FDRW): the fisher discriminant ratio as described by Lorena and Carvalho [LdC10] and used by Poh et al. [PB04] measures the separability of classes, here, genuine and imposter scores. The higher the separability, the higher is the biometric source performance. The FDR and the corresponding weights are given as

$$FDR_k = \frac{(\mu_k^G - \mu_k^I)^2}{(\sigma_k^G)^2 + (\sigma_k^I)^2}, \tag{4.6}$$

$$w_{k\_FDRW} = \frac{FDR_k}{\sum_{k=1}^{N} FDR_k}. \tag{4.7}$$

e) Brute force weighted (BFW): here, the weights are assigned by brute force search for optimal weights (weights that produces lowest EER) on the training data. This method is computationally expensive especially for higher order multi-biometrics. Therefore, only bi-modal biometric fusion were evaluated by BFW in this work.

f) Equal weighted (EQW): equal weighting assigns equal weights to all biometric sources under the assumption that all sources have the same contribution to the final fused biometric decision. This is usually used as a baseline and when no sufficient information (data) are available for the biometric sources in hand. EQW is equivalent to the basic sum/mean rule fusion.

The use of EERW, DPW and FDRW is a common practice for weighting biometric sources. However, more recent approaches just as the NCWW proved superiority over such approaches [CSN10]. Using brute force to assign weights has high computational expense and produces less generalized results as shown later in Section 4.3.2. The high performance of NCWW is however fragile as the NCW calculation depends on extrema values of comparison scores, which makes its performance very sensitive to outliers in training data.

4.3 Confident biometric source weighting

This section presents and evaluates a multi-biometric weighting approach that aims at capturing the confidence of each biometric source as well as its over all performance, while avoiding the influence of outliers in the training data.

4.3.1 Methodology

Two weighting algorithms are proposed here based on the properties of the genuine and imposter comparison scores distributions. First is the mean-to-extrema weighting (MEW) that depends on the mean values of the distributions with respect to their extremas. The second is the overlap deviation weighting (OLDW) that tries to avoid depending on unstable information such as distribution extrema (e.g. NCWW), and rather depends on more robust measures that represent both the confidence and the absolute performance of each biometric source.

g) Mean-to-extrema weighted (MEW): based on the assumption that a biometric source with low performance produces genuine score distribution that has a wide mean-to-min ranges and a wide mean-to-max imposter scores distribution range. The genuine mean-to-min range represents the difference between the mean of the genuine scores distribution and the minimum value (least correct) of the distribution. The same applies for the mean-to-max range in the imposter scores distribution.

The MEW is based on the width of the area between the mean of the imposter scores distribution and its maxima. It also considers the width of the area between the mean of the genuine scores distribution and its minima. This aims at focusing on the overlap area and its neighbor in both distributions. The MEW is formulated as

$$ME_k = (Max_k^I - \mu_k^I) + (\mu_k^G - Min_k^G), \tag{4.8}$$

$$w_{k\_MEW} = \frac{\frac{1}{ME_k}}{\sum_{k=1}^{N} \frac{1}{ME_k}}. \tag{4.9}$$

h) Overlap Deviation weighted (OLDW): this weighting approach is based on two assumptions, first is the inverse relation between the confidence of a biometric system and the standard deviation of the overlap area in its genuine-imposter scores distributions. The second assumption is the inverse relation between the EER value produced by a certain biometric source and its performance.

Overlap deviation tries to capture the properties of the overlap area between the imposter and genuine scores distributions without depending on singular extrema values. It also integrates the overall performance (FRR and FAR as an EER value) of the biometric source. Taking the standard deviation of this area aims at reducing the sensitivity to outliers in the data with respect to considering the width of the area. More importantly, this standard deviation is related to the confidence in a biometric source, as a biometric system producing less spread (in smaller range) genuine and imposter scores around their overlap region points out its decision confidence. Including the overall verification performance of the biometric source (EER) in the weighting process aims at creating a better generalized solution.

Given the imposter scores S_k^I, the genuine scores S_k^G, the EER and the score threshold at the equal error operating point T, the OLDW can be given as

$$OLD_k = \sigma(\{S_k^I \mid S \geq T\} \cup \{S_k^G \mid S < T\}) \times EER, \tag{4.10}$$

$$w_{k\_OLDW} = \frac{\frac{1}{OLD_k}}{\sum_{k=1}^{N} \frac{1}{OLD_k}}. \tag{4.11}$$

In the next Section 4.3.2, the experiment design is introduced. The performance achieved by the different weighting algorithms discussed will be presented.

4.3.2 Experimental setup

The proposed approaches for multi-biometric source weighting are general and can be applied to any number of biometric sources. However, the presented results focus on the case of bi-modal biometrics to investigate the performance away from high order complexities. Moreover, the performance of high order multi-biometric scenarios is also investigated.

Two multi-biometric scores databases were used to develop and evaluate the discussed solutions in order to assist the generalization capabilities of these solutions.

The first database is the XM2VTS [PB06, PBK10]. The Lausanne Protocols I (LP1) partition of the XM2VTS database was used in the experiment. This partition contains comparison scores produced by five face (F0 - F4) and three speech (S5 - S7) baseline experts. The evaluation and development sets defined by the authors were used in the performed experiments. The experiments here considered all possible pairs between face and speech matchers as well as the fusion of all matchers. The database contains 295 individuals, which results in 1000 genuine and 151,800 imposter scores. For more details about the XM2VTS score database, one can refer to the work of Poh et al. [PB06].

The second database used is the BSSR1 database [NIS]. The database contains comparison scores for left and right fingerprints (Fli and Fri) and two face matchers (Fc and Fg). BSSR1 multimodal database contains 517 genuine and 266,772 imposter scores. The experiments here considered all possible pairs between finger and face matchers as well as the fusion of all matchers. To evaluate the statistical performance of the proposed solutions, the database was split into three equal-sized partitions. Experiments were performed on all possible fold combinations, where one partition is used as evaluation set and the other two are used as a development set. The reported results are the averaged results of the three evaluation/development combinations.

Min-max normalization was used to bring comparison scores produced by different biometric sources to a comparable range [DON13]. Min-max normalized score $S'_{min-max}$ is given as

$$S'_{min-max} = \frac{S - min\{S_k\}}{max\{S_k\} - min\{S_k\}}, \tag{4.12}$$

where $min\{S_k\}$ and $max\{S_k\}$ are the minimum and maximum value of scores existing in the training data of the corresponding biometric source.

To produce the fused scores, the weighted-sum rule (linear combination) was used. The weighted-sum rule assigns each score value S_k with the weight of its source w_k. The weights w_k are calculated from the training data of each biometric source as discussed in Section 4.4.1. The fused score F by the weighted-sum rule for N score sources is given as

$$F_{weighted-sum} = \sum_{k=1}^{N} w_k S_k, k = \{1, \ldots, N\}. \tag{4.13}$$

The performance of the fusion process under different weighting approaches is evaluated under verification scenario and presented as EER values and as ROC curves.

For each of the databases, all bi-modal combinations are evaluated along with the overall fusion of all available sources. As expected, the results show the advantage of multi-biometrics on the verification performance.

4.3.3 Results

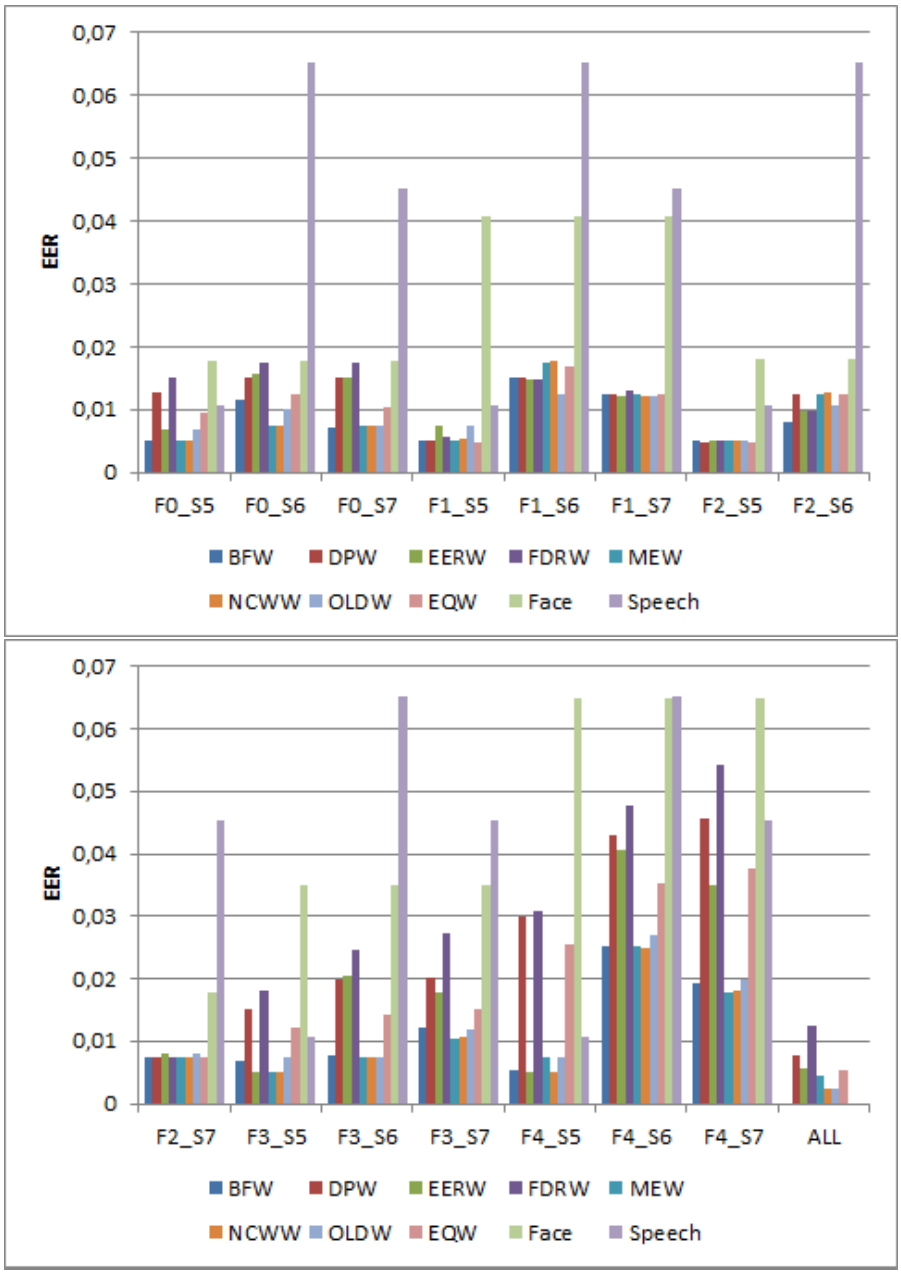


Figure 4.1: EERs achieved on the XM2VTS database: the rates shown here represent all possible bi-modal combinations of face matchers (F0 - F4) and speech matchers (S5 - S7) in the XM2VTS database and the results achieved by the fusion of all eight available sources.

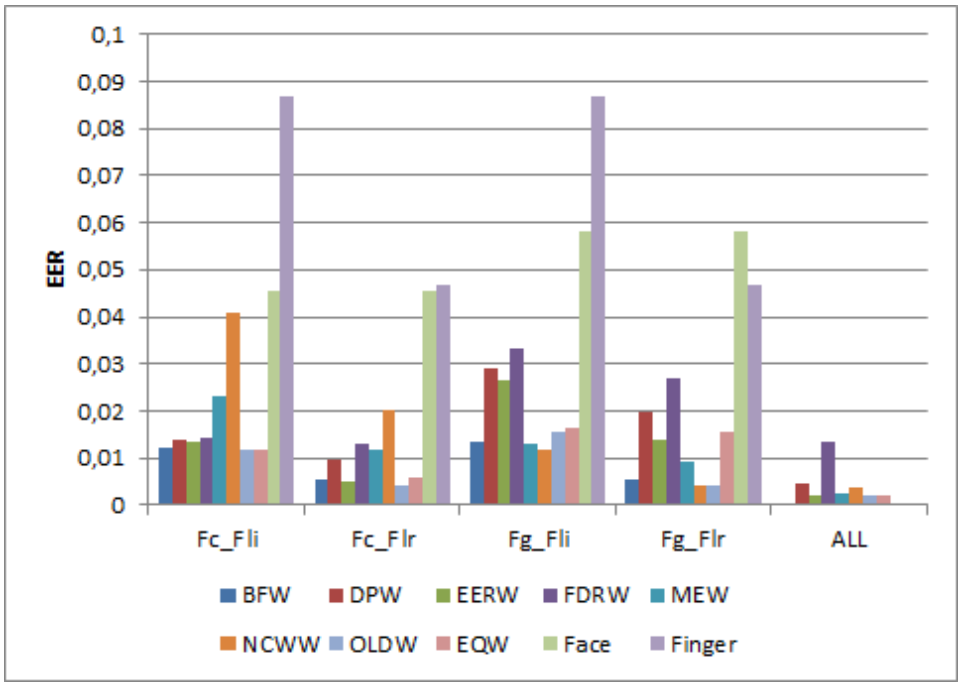


Figure 4.2: EERs achieved on the BSSR1 database: the rates shown here are for all possible bi-modal combinations of face matchers (Fc and Fg) and finger matchers (Fli - Flr) in the BSSR1 database and the results achieved by the fusion of all four available sources.

The EER values obtained from the XM2VTS database are shown in Figure 4.1. One can notice the high performance of NCWW and OLDW both scoring 0.25% EER for the overall fusion evaluation followed by the MEW with 0.46% EER and far from the commonly used DPW with 0.75% EER. In the bi-modal evaluation, NCWW showed high performance in many combinations closely followed by a stable performance by the OLDW. It must be noticed that in some scenarios such as in $F1\_S6$ and $F2\_S6$ NCWW performed worse than most approaches while OLDW sustained stable high performance.

EER values obtained from different approaches using the BSSR1 database are shown in Figure 4.2. The figure shows the superiority of the OLDW approach in most cases with stable performance compared to the NCWW approach. In the overall fusion evaluation, the OLDW scored the best performance at 0.21% EER followed by EERW, EQW and MEW while the NCWW scored 0.37% EER. The fluctuation in the NCWW performance, with respect to that of the OLDW, may be related to its dependence on extrema values that are more vulnerable to outliers than the measures used to calculate the OLDW.

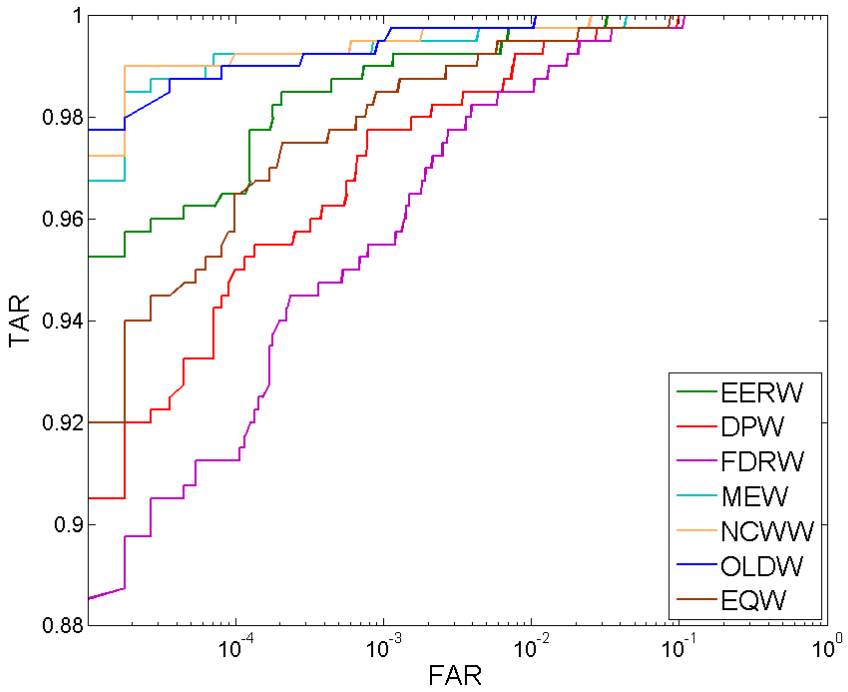


Figure 4.3: ROC curves achieved on the XM2VTS database: the curves shown here represent the performance of the fusion of all eight (five face and three speech) available sources using different weighting approaches.

Results of evaluation over the XM2VTS database using different weighting approaches to fuse all available sources (five face and three speech) are also shown as ROC curves (Figure 4.3) to investigate the performance under different operational points. On very low FAR, the proposed OLDW performs the best. While the FAR values get higher, the lowest FRR is achieved by the NCWW and the proposed OLDW and MEW approaches.

The ROC curves achieved on the BSSR1 database are shown in Figure 4.4. These curves are graphically averaged curves over the three testing folds of the database similarly to vertical averaging discussed in [PMB07]. One can notice the superiority of the OLDW performance, especially at low FAR, followed by the MEW and the NCWW.

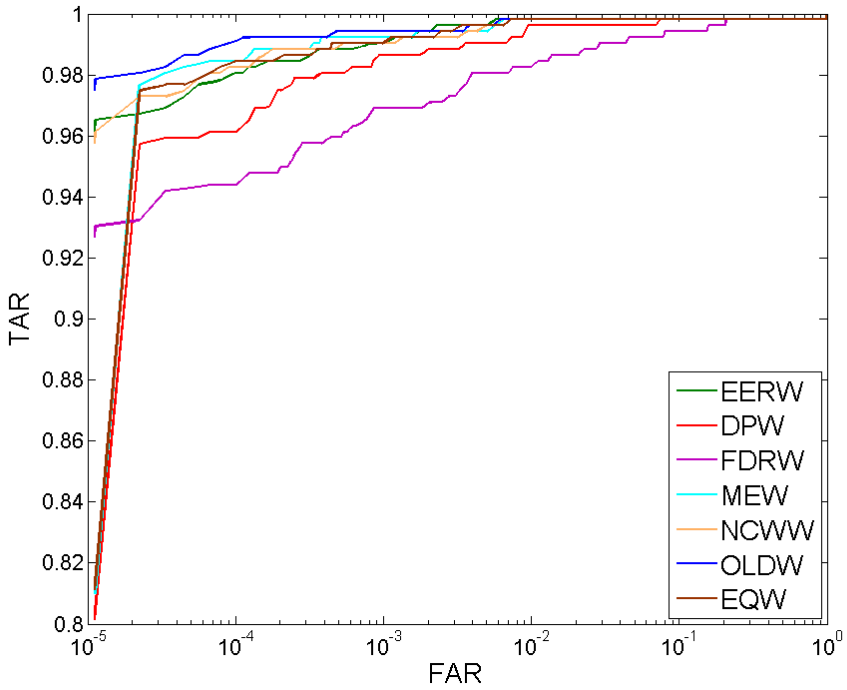


Figure 4.4: ROC curves achieved on the BSSR1 database: the curves shown here represent the performance of the fusion of all four available sources (two face matchers and two fingerprint matchers) using different weighting approaches.

4.4 CMC curve properties and biometric source weighting

This section presents and evaluates a multi-biometric weighting approach that captures the identification performance of each biometric source by analyzing properties of the CMC curve achieved by these sources.

4.4.1 Methodology

A CMC curve represents the performance of a biometric system within a close-set identification scenario. The horizontal axis is the number of the top ranks considered and the vertical access is the probability of the correct identity (the identity of a captured subject) being within the top ranks considered. This cumulative probability at a certain rank r will be referred to as the true identification rate at that rank $TIR(r)$. As a performance measure, the properties of the CMC curve can influence the relative importance of single biometric sources within the process of multi-biometric fusion. The weighting approaches proposed here and based on CMC curve properties aim at being generalized and robust to outliers as the shape of such performance measure (CMC) is less affected by outliers in the data.

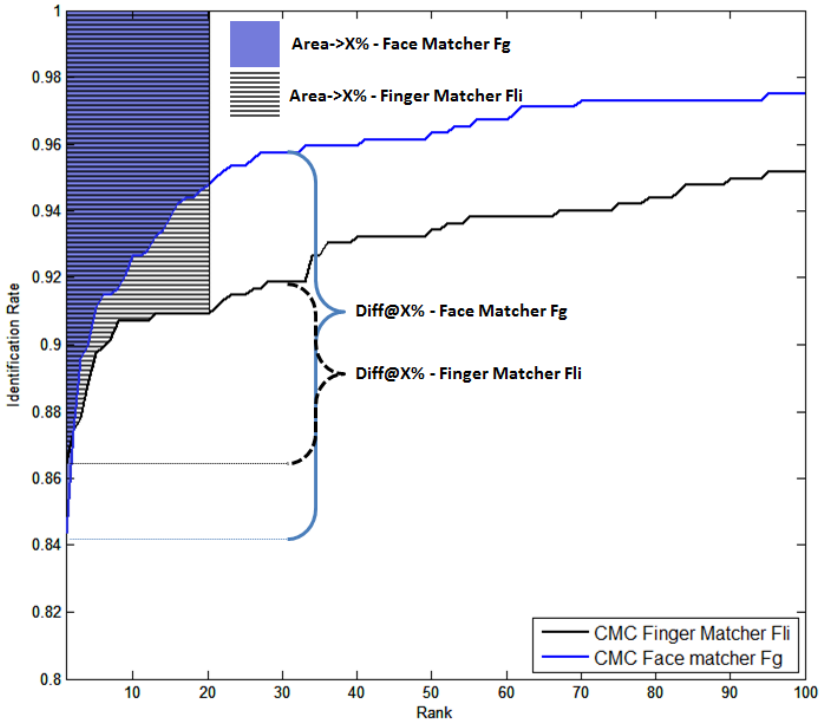


Figure 4.5: CMC curves of two different matchers (Finger Fli and Face Fg) of the BSSR1 database. The difference in the limited area above the CMC curves of the two different matchers is shown along with the difference in the CMC elevation between both matchers at a fixed limit.

Three properties of the CMC curve are considered in this work. First is the *limited area above the CMC curve*, which is the area above the CMC curve between rank one and a given rank. As the CMC curve is affected by the number of identities L in the search domain, the high limit of the calculated area is considered as a percentage X of the total L and not a fixed rank. One can notice in Figure 4.5 that a better performing biometric source produces a CMC curve with lower area above the curves. Knowing that the $TIR(r)$ is the probability that the correct identity of a given subject is among the highest r achieved scores within a closed set biometric identification system, the area above the CMC curve is discretized and is calculated as

$$\text{Area}{\rightarrow}\text{X}\%_k = \sum_{r=0}^{r=\lfloor X*L/100 \rfloor} TIR(r), \tag{4.14}$$

and therfore the weights can be given as

$$w_{k\_\text{Area}{\rightarrow}\text{X}\%} = \frac{\frac{1}{\text{Area}{\rightarrow}\text{X}\%_k}}{\sum_{k=1}^{N} \frac{1}{\text{Area}{\rightarrow}\text{X}\%_k}}. \tag{4.15}$$

In the evaluation, three values of X where considered ($X = 2\%$, 4% and 20%) and the weights by the area above the CMC curve with these limits are noted as Area→2%, Area→4%, and Area→20%.

The second property of CMC curves considered here is the *elevation of the CMC curve*. The elevation can be given by the difference in the TIR between rank one and a certain rank limit. This rank limit is given as a percentage X of the total number of identities L in the search domain. The assumption that a better performing biometric source has a higher elevation (improve faster when considering more ranks) is clear in Figure 4.5. The elevation is formulated as

$$\text{Diff@X\%}_k = TIR(\lfloor X * L/100 \rfloor) - TIR(0). \tag{4.16}$$

As the elevation is in direct proportion to the biometric source performance, the weight based on the elevation is calculated by

$$w_{k\_\text{Diff@X\%}} = \frac{\text{Diff@X\%}_k}{\sum_{k=1}^{N} \text{Diff@X\%}_k}. \tag{4.17}$$

Two values of X were considered ($X = 2$ and 4%) and the weighting by the elevation of the CMC curve with these limits are noted as Diff@2% and Diff@4%.

The third property of the CMC curve considered in this work is the *true identification rate (TIR) within a number of top ranks*. At a given rank, which is defined by a percentage X of the number of identities L in the search domain, a better performing biometric source achieves higher TIR. Here, the false identification rate $FIR(r)$ at a certain rank r was considered as a compliment to the TIR as $FIR(r) = 1 - TIR(r)$. Thus, the related FIR value at that rank can be given as

$$\text{FIR@X\%}_k = FIR(\lfloor X * L/100 \rfloor) = 1 - TIR(\lfloor X * L/100 \rfloor), \tag{4.18}$$

and the weights here are calculated by

$$w_{k\_\text{FIR@X\%}} = \frac{\frac{1}{\text{FIR@X\%}_k}}{\sum_{k=1}^{N} \frac{1}{\text{FIR@X\%}_k}}. \tag{4.19}$$

Three variations of these measures were tested in this work and are taken at the first rank (R1), $X = 2\%$ and $X = 4\%$ and are noted by TIR@R1, TIR@2%, and TIR@4% (TIR@X%$_k = TIR(\lfloor X * L/100 \rfloor)$).

In the next section, the evaluation results of the proposed weighting solutions are presented, discussed, and compared to baseline solutions.

4.4.2 Experimental setup

The proposed approaches for multi-biometric source weighting are general and can be applied to any number of biometric sources. However, as in Section 4.3.2. the presented results focus on the case of bi-modal biometrics to investigate the performance away from high order complexities. Moreover, the performance of high order multi-biometric scenarios is also investigated when the results are discriminant.

The database used to develop and evaluate the proposed solution is the Biometric Scores Set BSSR1 - multi-modal database [NIS]. This database is described in details in Section 4.3.2.

Min-max normalization was used to bring comparison scores produced by different biometric sources to a comparable range. Min-max normalized score is given as in Equation 4.12. To produce the fused scores, the

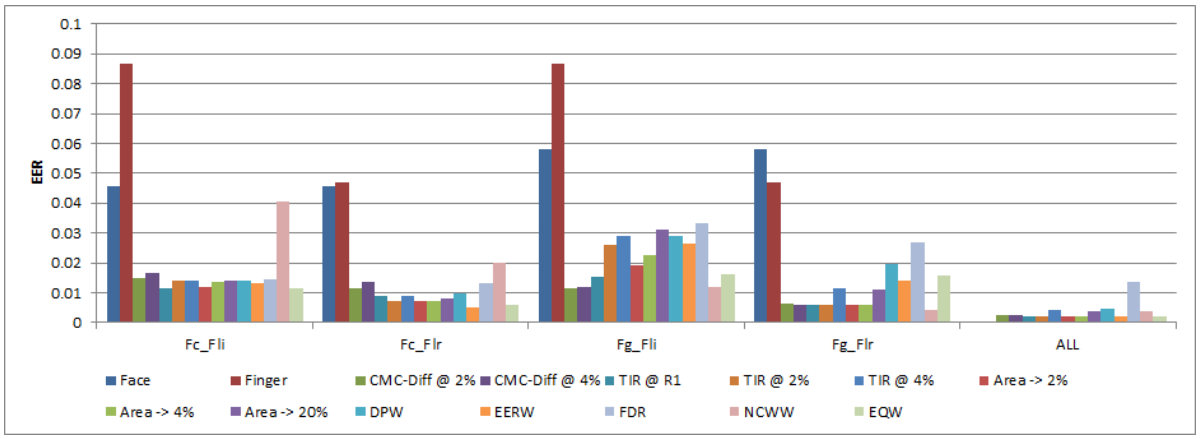


Figure 4.6: EERs achieved on the BSSR1 database: the rates shown here are for all the possible bi-modal combinations of face matchers (Fc and Fg) and finger matchers (Fli - Flr) in the BSSR1 database and the results achieved by the fusion of all four available sources (two face matchers and two fingerprint matchers).

weighted-sum rule (linear combination) was used. The weighted-sum rule assigns each score value S_k with the weight of its source w_k. The weights w_k are calculated from the training data of each biometric source as discussed in Section 4.4.1. The fused score by the weighted-sum rule is given in Equation 4.13.

The performance of the fusion process under different weighting approaches is evaluated under verification scenario and presented as EER values and as ROC curves. The performance under the identification scenario was evaluated and the achieved performance is presented as CMC curves.

4.4.3 Results

Figure 4.6 shows the EER values obtained from the BSSR1 database using different proposed and baseline weighting approaches. The EER values are presented for all possible bi-modal combinations as well as for the fusion of all available sources (two face and two fingerprint matchers). One can notice the relatively stable and high performance of the weighting approaches based on the CMC properties. The lowest EER for fusing all available sources was achieved by the approaches based on the area above CMC (Area$\rightarrow$2%) with 0.199% EER followed by the Area$\rightarrow$4% and the CMC identification rate based approaches TIR@R1 and TIR@2% with all three scoring an EER below 0.205%. The closest performance from the baseline approaches was the EER based weighting with an EER of 0.221%.

The ROC curves achieved on all bi-modal combinations are shown in Figures 4.8a, 4.8b, 4.8c and 4.8d. The ROC curves achieved by the CMC based weighting techniques for fusing all the four biometric sources are presented in Figure 4.8e. The best of the Figure 4.8e curves are compared to the curves achieved by the baseline solutions in Figure 4.8f.

In the bi-modal combinations ROC curves in Figures 4.8a, 4.8b, 4.8c and 4.8d, it can be noticed that the highest performances at a very low FAR were achieved by approaches based on the area above CMC curve and identification rates at first rank. At higher FAR, weighting approaches based on elevation in the CMC curve

becomes more competitive scoring higher true acceptance rate (TAR) than approaches based on the area above CMC curve in some operation areas.

In the case of the fusion of all four biometric sources, Figure 4.8e shows that the highest TAR rates at very low FAR were achieved by the Area→2%, Area→4% and TIR@R1 approaches. The approaches based on elevation in the CMC curve scores relatively high TAR on higher FAR values.

When compared to the baseline biometric source weighting approaches, the superiority of CMC properties based weighting is clear, see Figure 4.8f. This is more obvious on very low FAR values especially with the Area→2%, Area→4% and TIR@R1 approaches.

Performance under the identification scenario was evaluated and presented as CMC curves for bi-modal combinations, see Figure 4.7. Fusion of all the four biometric sources achieved very high and incomparable performance due to the limited size of the database and therefore, this case was not presented. The CMC curve elevation based weighting scored the highest first rank identification rate in two of the four combinations as shown in Figures 4.8a and 4.8b and coming close to the best in the other two combinations, see Figures 4.8c and 4.8d. Approaches based on the area above CMC curve and weighting based on first rank identification rate also had a satisfying and stable performance over all four combinations. Some baseline approaches achieved good performance such as the EERW but felt short behind CMC curve based weighting solutions. The NCWW weighting also achieved good results in some combinations but with fluctuating overall performance.

One must notice the different performance behavior of weighting approaches under both identification and verification scenario. A clear example is the high performance of the CMC curve elevation based approaches under the identification scenario with respect to their performance under the verification scenario.

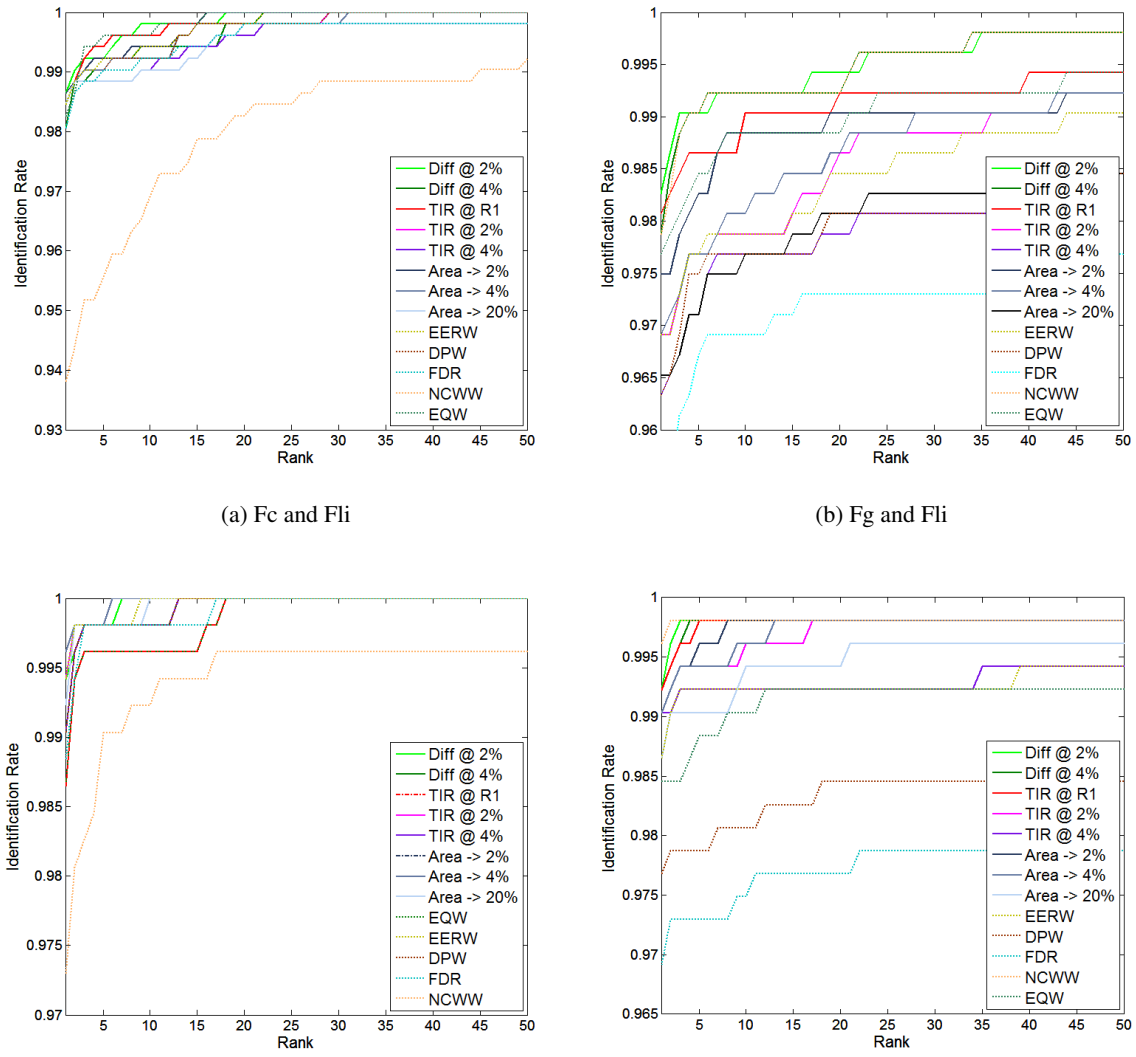
4.5 Summary

This chapter discussed optimizing the widely implemented weighted-sum fusion rule in multi-biometric score-level fusion. Two aspects were in focus, first is introducing a weighting approach that considers both the confidence and the absolute performance of biometric sources. Secondly, designing biometric source weighting scheme based on identification performance measures and analyzing its effect on the multi-biometric performance, especially under the identification scenario.

Score-level multi-biometric fusion is usually performed using the weighted-sum rule where each source is assigned a weight to control its influence in the final decision. Different weights have been proposed in the literature with a focus on representing the relative performance of the different sources. However, no emphasis on the confidence of the decisions of single sources was introduced in these weights. On the other hand, most of the previously proposed weights were based on the biometric verification performance measures and did not consider identification performance that relies on a large number of comparisons.

At first, this chapter introduced a weighting approach that aimed at combining indication of the performance and the source confidence. A confident source was seen as a source that provides a more distinct mapping between its possible verification decision and the comparison scores it produces. To achieve this, the approach considered the standard deviation of the overlapped area between genuine and imposter scores distributions to represent the source confidence. It also considered the overall performance (as EER) achieved by each biometric source. To prove the validity of the proposed weighting scheme, an evaluation was conducted on the XM2VTS and the BSSR1 databases. These experiments showed a consistent high multi-biometric verification performance in comparison to state-of-the-art approaches, especially at low FAR operation points.

Looking into biometric sources weighting based on their identification performances, this chapter proposed a number weighting techniques based on the properties of the CMC curves produced by these sources. These



(a) Fc and Fli

(b) Fg and Fli

(c) Fc and Flr

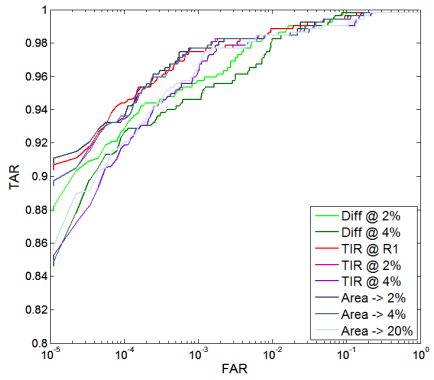
(d) Fg and Flr

Figure 4.7: CMC curves achieved on the BSSR1 database: the curves shown here are for all the possible bi-modal combinations of face matchers (Fc and Fg) and finger matchers (Fli - Flr) in the BSSR1 database. The Figures also represent a comparison between different CMC-based weighting approaches and baseline approaches. a) Fc and Fli, b) Fg and Fli, c) Fc and Flr. d) Fg and Flr.

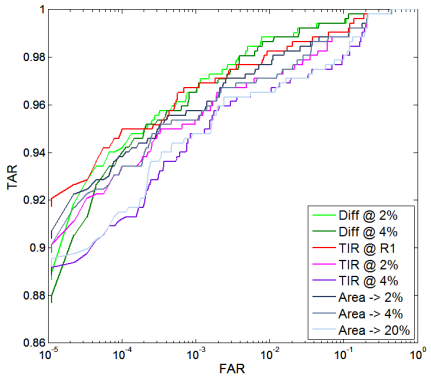
weights looked at the limited area above the CMC curve, the rapidity of improvement in the cumulative identification rate when moving up the identification ranks, and the cumulative identification rate at certain ranks. These approaches were evaluated under the verification and identification scenarios along with a number of state-of-

the-art solutions. The proposed weights based on the rate of improvement in the cumulative identification rate performed best under the identification scenario, which interestingly was not the case under the verification scenario. Weights based on the cumulative identification rate at fixed ranks and the area above the curve performed best under the verification scenario, outperforming the baseline solutions. This was based on an evaluation executed on the BSSR1 database.

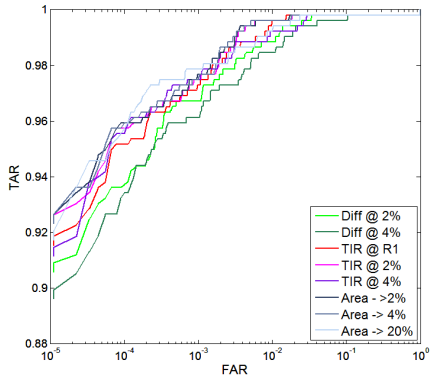
This chapter presented an answer to the *RQ3* by proposing and validating a multi-biometric source weighting approach that captures the absolute performance of a biometric source, as well as its confidence. It also responded to the *RQ4* by proposing a number of weighting approaches based on the biometric identification performance metric, the CMC curve, and analyzing the relative effect of these weights on the identification and verification scenarios. The next chapter 5 is also concerned with the optimization of the fusion process. However, it will focus on proposing new supplementary information derived from comparison scores to further optimize this process.



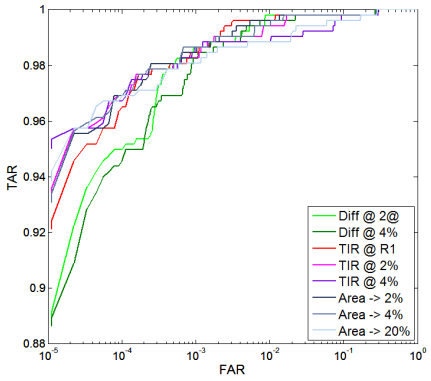
(a) Fc and Fli.



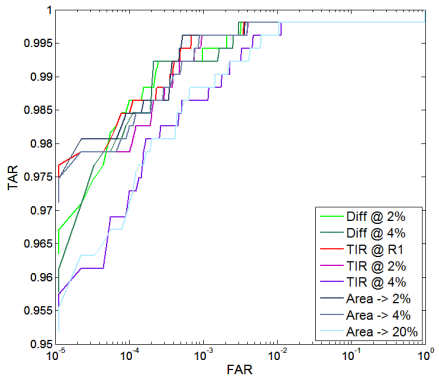
(b) Fg and Fli.



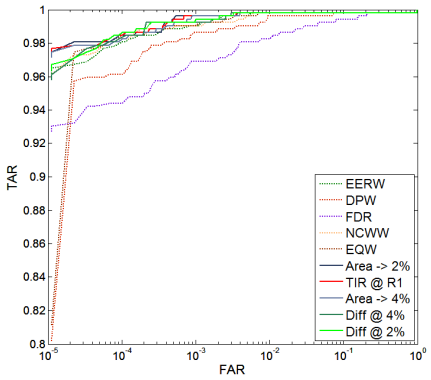
(c) Fc and Flr.



(d) Fg and Flr.



(e) All sources (Fc, Fg, Fli, and Flr).



(f) All sources (Fc, Fg, Fli, and Flr) vs. baseline.

Figure 4.8: ROC curves achieved on the BSSR1 database: all possible bi-modal combinations of face matchers (Fc and Fg) and finger matchers (Fli - Flr) in the BSSR1 database and the the fusion of all four available sources. Figures (a,b,c,d,e) shows a comparison between different CMC-based weighting approaches and Figure (f) represents a comparison to baseline approaches.

5 Integrating supplementary information

Last chapter was concerned with fusion optimization by focusing on multi-biometric source weighting. This chapter also aims at optimizing the fusion process. However, this is achieved through introducing supplementary information into the multi-biometric fusion process. Beside comparison scores from multiple biometric sources, this chapter proposes two additional evidences on the genuinity of a comparison. The first is based on the relation of the fused scores to other comparisons scores and the second considers the levels of decision coherence of the different sources coherence. This chapter is based on the published works [DO14, DN15, DRBK17, DABK17].

5.1 Introduction

Multi-biometrics aims at building more accurate unified biometric decisions based on the information provided by multiple biometric sources. As discussed earlier, information fusion is used to optimize the process of creating this unified decision. In previous works dealing with score-level multi-biometric fusion, the scores of different biometric sources belonging to the comparison of interest are used to create the fused score. Some works proposed to include the quality of the captured samples in the fusion process. However, these quality measures require the access to such samples as well as a well-defined quality measures that correspond to the biometric performance. This would be dependent on the performance of the quality estimators and makes the system less integratable as it requires access to raw data that might be restricted due to privacy and security limitations. The novelty in this chapter focuses on defining and integrating two new forms of supplementary information that consider the relations between the scores produced by the different biometric sources after the biometric comparison process rather than the raw data.

The first proposed supplementary information aims at integrating the relation of the fused scores to other comparisons within a 1:N comparison. This is performed by considering the neighbors distance ratio in the ranked comparisons set within a classification-based fusion approach. A further improvement is made by considering weighting the biometric sources to produce more informative neighbors distance ratios.

Keeping the open-set identification scenario in mind, this work tries to use the information provided by the ranked set of comparisons to perform more accurate verification of the top rank. The assumption here is that a genuine top rank comparison has a lower distance ratio to its rank neighbors than that of an imposter comparison. This information are represented as neighbors distance ratios and are integrated into a classification-based fusion approach using SVM. An enhanced solution is also proposed by considering the relative influence of the different biometric sources in the neighbors distance ratio assignment process.

The proposed fusion technique is evaluated over the biometric scores set BSSR1 - multimodal database (BSSR1) [NIS]. A number of previously proposed baseline fusion approaches were evaluated including state-of-the-art combination rules and the use of SVM without consideration of the neighbor distance ratio. The proposed technique proved to outperform the baseline solution and the results are presented as equal error rates (EER) and receiver operating characteristics (ROC) curves.

The second form of supplementary information aims at embedding score coherence information in the fusion process to further enhance the multi-biometric performance. This is based on the assumption that a minority of

biometric sources, which point out towards a different decision than the majority, might have faulty conclusions and should be given relatively smaller role in the final decision.

The use of the multi-biometrics scores coherence as a supplementary source of information in the fusion process is based on the assumption that, if most of the biometric sources point out to a certain decision (mainstream), the smaller number of sources pointing elsewhere might be misinformed (e.g. due to noisy captures) and thus should play a smaller role in the final decision. A coherence measure is defined and integrated in the fusion process as a dynamic weight along with a conventional static source weighting approach.

The proposed fusion technique is evaluated on the BioSecure multi-modal biometric database [OFA*10]. Different versions of the database were created by adding blur noise to a certain percentage of the raw data, to create a more realistic scenario. The enhanced performance induced by including the coherence information within a dynamic weighting scheme in comparison to the baseline solution was shown by the reduction of the equal error rate by 45% to 85% over the different test scenarios and proved to maintain high performance when dealing with noisy data. More importantly, in the scenarios where noisy data is involved, including the coherence information limited the effect of noise on the overall performance.

The next section contains a short overview of related works motivating and leading to the presented approach. Sections 5.3.5 and 5.4.5 present the two proposed supplementary information along with experimental comparisons to baseline solutions. A final discussion of the chapter contributions is presented in Section 5.5.

5.2 Related work

Score-level biometric fusion techniques can be categorized into two main groups, combination-based and classification-based fusion. Combination-based fusion consists of simple operations performed on the normalized scores of different biometric sources. These operations produce a combined score that is used to build a biometric decision. One of the most used combination rules is the weighted-sum rule, where each biometric source is assigned a relative weight that optimizes the source effect on the final fused decision. The weights are related to the performance metrics of the biometric sources, a comparative study of biometric source weighting is presented by Chia et al. [CSN10] and later extended by Damer et al. [DON14a]. Classification-based fusion views the biometric scores of a certain comparison as a feature vector. A classifier is trained to classify these vectors optimally into genuine or imposter comparisons. Different types of classifiers were used to perform multi-biometric fusion, some of these are support vector machines (SVM) [SVN07, GV00, DO14], neural networks [Als10], and the likelihood ratio methods [NCDJ08].

Conventionally, score-level multi-biometric fusion exclusively uses the biometric comparison scores provided by the fused biometric sources and general information about those sources (e.g. weights). Previous works extended this concept to include additional supplementary information. Sample quality information related to each biometric comparison was integrated into the fusion process, resulting in accuracy gain [NCJD06, PK08, PMK09], but requires appropriate quality definitions and access to raw data.

More advanced approaches of multi-biometric fusion considered dynamic weights that adapt to the comparison set in hand. Hui et al. proposed a dynamic weighting approach for multi-biometric fuzzy-logic based fusion [HMM07]. The dynamic weights took into account the variations during data acquisition as supplementary information (e.g. lighting, noise and user-device interactions), which requires previous knowledge or reliable detection of these variations. Other works applied dynamic weights based on capture quality and scenario on a feature level fusion process [WWG*07, YLHZ12].

This chapter proposes two types of supplementary information that do not require the access to any information prior to the biometric comparison, therefore depend only on the multi-biometric scores. Therefore, the

proposed approaches are inherently more integratable and robust to errors resulting from environment and quality estimators.

5.3 Biometric neighbors distance ratio

This section presents and evaluates the first supplementary information form presented in this chapter, the biometric neighbors distance ratio (NDR). A number of baseline solutions used to benchmark the performance are also discussed. Most of the baseline approaches were introduced previously in Section 4.2 and are reintroduced here.

5.3.1 Baseline solution

A number of baseline solutions are presented here to build a reference for the performance evaluation presented in the next Section 5.3.5. The first baseline solution aims at direct comparison by using SVM-based approach for fusion without using NDR information. Two other solutions based on the widely used weighted-sum approach are also discussed, one utilizes the EER as a source performance measure while the second uses the non-confidence width (NCW).

The conventional SVM baseline approach takes the biometric comparison scores of different sources $\{S_1,\ldots,S_n\}$ as a feature vector. The SVM is trained to classify this feature vector into genuine or imposter classes and report the resulted decision function value as the fused score. The SVM used here also uses similar configuration to the proposed approach with RBF as a kernel function.

The two other baseline approaches are based on the weighted-sum combination rule that assigns each score value S_k with the weight of its source w_k to produce the fused score. The weights w_k are calculated from the training data of each biometric source. The fused score F by the weighted-sum rule for N score sources is given as

$$F_{weighted-sum} = \sum_{k=1}^{N} w_k S_k, k = \{1,\ldots,N\}. \tag{5.1}$$

The weights used here are based on either EER or NCW values. These weights are discussed in Section 4.2 and are briefly reintroduced here. The EER weighting (EERW) is based on the EER value, which is the common value of the false acceptance rate (FAR) and the false rejection rate (FRR) at the operational point where both FAR and FRR are equal. EER weighting was used to linearly combine biometric scores in the work of Jain et al. [JNR05]. The EER is inverse proportional to the performance of the biometric source. Therefore, for a multi-biometric system that combines N biometric source, the EER weight for a biometric source k is given by

$$w_{k\_EERW} = \frac{\frac{1}{EER_k}}{\sum_{k=1}^{N}\frac{1}{EER_k}}. \tag{5.2}$$

The non-confidence width weight (NCWW) was proposed by Chia et al. [CSN10] to weight biometric sources for score-level multi-biometric fusion. NCW corresponds to the width of the overlap area between the genuine and imposter scores distributions. Given that Max_k^I is the maximum imposter score and Min_k^G is the minimum genuine score, NCW is given by

$$NCW_k = Max_k^I - Min_k^G, \tag{5.3}$$

as the NCW is inverse proportional to the biometric source performance, the weights based on the NCW is given as

$$w_{k\_NCWW} = \frac{\frac{1}{NCW_k}}{\sum_{k=1}^{N} \frac{1}{NCW_k}}. \tag{5.4}$$

5.3.2 Neighbors distance ratio

The main assumption that builds the bases of the proposed solution in this work is anchored on NDR. Given a rank set of comparison scores that represent a $1:N$ comparison, NDR is defined as the ratio between one score in this set and a score of a higher rank (neighbor distance). A similar concept was previously used in the literature to match interest key point descriptors in images [MS05].

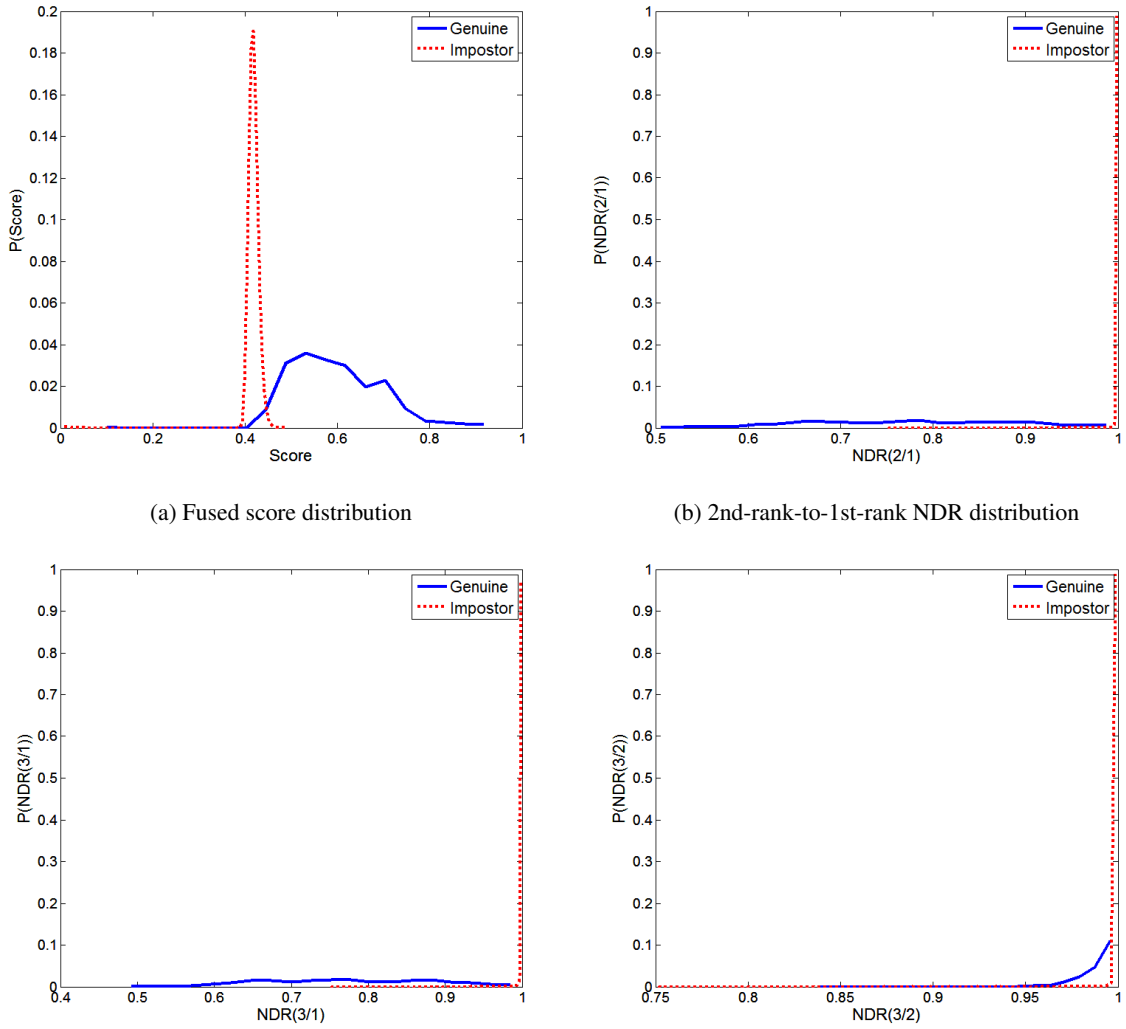
Looking into the NDR from the biometric perspective, the inverse ratio between a genuine similarity score and the next highest score (within a ranked $1:N$ comparison) is assumed to be lower than this ratio between an imposter score and the next highest score. In different words, a genuine score of a certain subject is relatively higher and distanced from the set of a clustered set of imposter scores produced by the same subject. This relative (to neighbor ranks) difference is not considered in conventional biometric verification where only the absolute value of the genuine or imposter comparison score is considered. A realistic genuine-imposter distributions of NDR values are shown and discussed later on.

The proposed solution in this work aims at considering both, the scores absolute values and the relative distances to higher ranks in order to perform more accurate biometric verification. To achieve that, a classification-based fusion approach based on SVM was used. In classification-based multi-biometric fusion, the fusion process is viewed as a binary classification problem that aims to separate between two classes, genuine and imposter.

Support vector machines [Vap95] is a statistical learning technique often used to learn binary classifiers, i.e. to learn how to separate two classes using information gained from known examples (training data). Classical learning techniques, such as neural networks (NN), focused on minimizing the empirical error (error on the training set). This approach is commonly referred to as empirical risk minimization (ERM). However, the SVM follows the structural risk minimization (SRM) instead of the ERM approach. The SRM insures a high generalization performance as it tries to minimize the upper bound of the generalization error. In simple words, SVM tries to build a class-separation surface in the feature space that is optimized in a manner which considers generalized unknown data.

In order to map the input data space into a feature space where the data is linearly separable, SVM uses kernel functions. In general, these functions help enhance the discrimination power. In this work, the radial basis function (RBF) is used as a kernel function as it proved to outperform linear kernels when dealing with low dimensional space [SZL*11], such as the problem dealt with in this work.

The feature vector considered for the SVM fusion process consisted of two concatenated parts. First is the set of comparison scores achieved by the N biometric sources, $\{S_1, \ldots, S_n\}$. The second part is a set of NDR values produced by the ranked list of the fused scores of the biometric sources. To produce these values, each biometric comparison (containing N scores) was fused using the simple sum rule. The resulted fused scores set is ranked and the NDR values are calculated. The considered NDR values were the 2nd-rank-to-1st-rank, 3rd-rank-to-1st-rank, and the 3rd-rank-to-2nd-rank values. The distribution of these values over imposter and genuine comparisons of the BSSR1 [NIS] fingerprint (Fli) and face (Fc) matchers (fused) are shown in Figure 5.1 where the high discrimination between imposter and genuine NDR values is clear. More details about the BSSR1 database are presented in Section 5.3.4.



(a) Fused score distribution

(b) 2nd-rank-to-1st-rank NDR distribution

(c) 3rd-rank-to-1nd-rank NDR distribution

(d) 3rd-rank-to-2nd-rank NDR distribution

Figure 5.1: the NDR values distributions for the fused BSSR1 [NIS] fingerprint (Fli) and face (Fc) matchers. Distributions points out that the NDR values can clearly discriminate between genuine and imposter comparisons.

The NDR values were concatenated with the comparison score values of the comparison resulting in a feature vector of the size $N+3$. The SVM classifies the input feature vector and the resulted decision function value (the signed distance to the margin) is considered as the final fused score. An overall look on the proposed method is presented in Figure 5.2. This solution will be referred to as SVM-NDR.

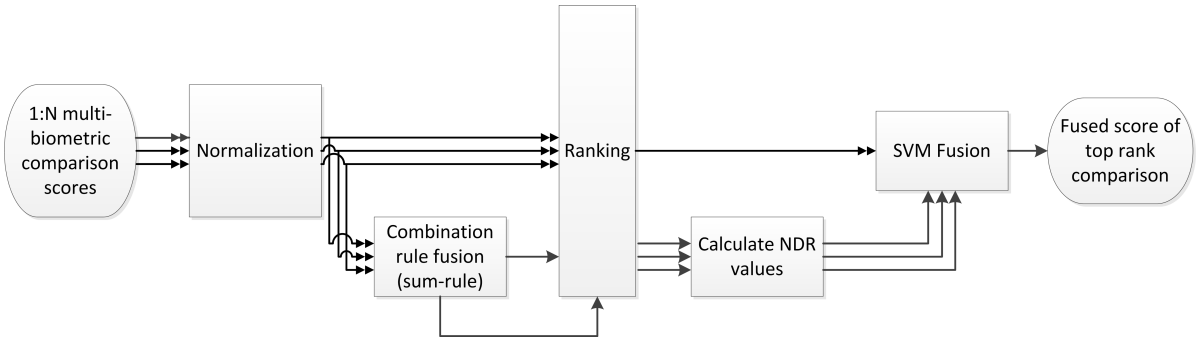


Figure 5.2: an overview of the proposed basic NDR solution. The input scores of a 1:N comparison is fused by simple sum combination rule and ranked based on the resulted scores. The NDR values based on this ranking is concatenated with the original scores of the comparison to be verified. The concatenated vector is fed into the SVM to create a final fused score.

5.3.3 Weighted NDR integration

The idea of integrating the NDR values within a classification based fusion solution is extended here by introducing information about the performance of the different biometric sources into the NDR calculation process. This will help in producing more accurate initial ranking and more informative NDR values. The performance information were included as biometric source weights calculated based on the OLDW [DON14a].

The proposed modification is based on providing more accurate initial ranking to calculate NDR values. This is achieved by using OLDW [DON14a] approach (presented in Section 4.3.1) to assign relative weights to different biometric sources to influence their effect on the overall initial ranking, and thus the accuracy of the NDR values. The weighted biometric scores also affect the values of the NDR as the initially fused scores are also fused by a weighted-sum rule.

As in the basic proposed solution, the feature vector considered for the SVM fusion process consisted of two concatenated parts, the initial comparison scores of different sources N and the NDR values based on the initial weighted fusion. Here, three NDR values were considered, 2nd-rank-to-1st-rank, 3rd-rank-to-1st-rank, and the 3rd-rank-to-2nd-rank. This will result in a feature vector of size $N+3$. The SVM classifies the input feature vector and the resulting decision function value (the signed distance to the margin) is considered as the final fused score. An overall look on the weighted NDR integration method is presented in Figure 5.3. This solution will be referred to as Weighted SVM-NDR.

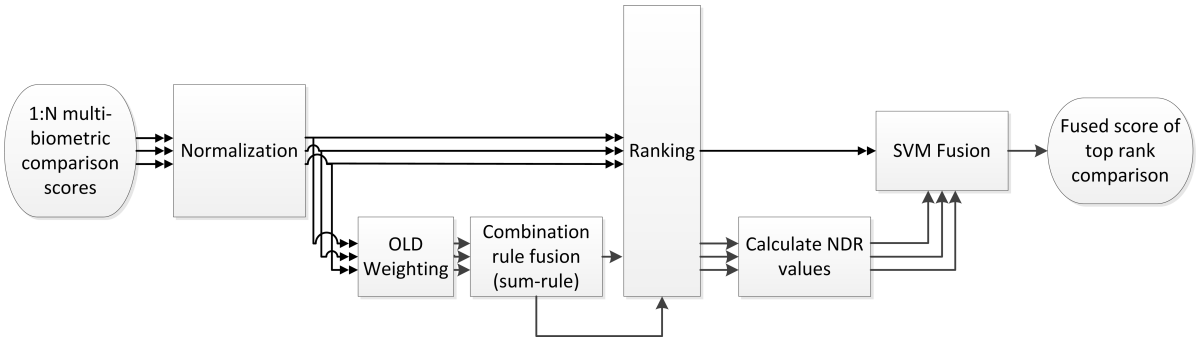


Figure 5.3: an overview of the proposed weighted NDR solution. The input scores of an 1:N comparison is weighted (OLDW) and fused by simple sum combination rule then ranked based on the resulted scores. The NDR values based on this ranking is concatenated with the original scores of the comparison to be verified. The concatenated vector is fed into the SVM to create a final fused score.

5.3.4 Experimental setup

The database used to develop and evaluate the proposed solution is the biometric scores set BSSR1 - multimodal database (BSSR1) [NIS]. The database contains comparison scores for left and right fingerprints (Fli and Fri) and two face matchers (Fc and Fg). The BSSR1 database contains 517 genuine and 266,772 imposter scores. The experiments here considered all possible pairs between finger and face matchers as well as the fusion of all matchers. To evaluate the statistical performance of the proposed solutions, the database was split into three equal-sized partitions. Experiments were performed on all possible fold combinations, where one partition is used as an evaluation set and the other two are used as a development set. All the reported results are the averaged results of the three evaluation/development combinations.

Min-max normalization was used to bring comparison scores produced by different biometric sources to a comparable range. Min-max normalized score is given as in Equation 4.12.

To train and test the proposed approach, every possible open-set identification scenario that can occur in the database was simulated. To do that, the comparisons in the database were split into separated 1:N comparison sets. Each comparison in these comparison sets was fused using the sum rule fusion then ranked according to the resulted fused scores. The three considered NDR values were calculated for each entry in the ranked comparison sets, except the last two ranks, as the second and third rank to these entries does not exist and thus the NDR values cannot be calculated. The resulted NDR values of each comparison are concatenated with the original scores of the comparison to create the final feature vector for that comparison. The resulted feature vectors are passed along with their genuine/imposter labels to train the SVM classifier in the training mode.

For evaluation, similar concatenated feature vectors are created from the testing data. These features are evaluated by the trained SVM classifier to produce a final fused score from each comparison.

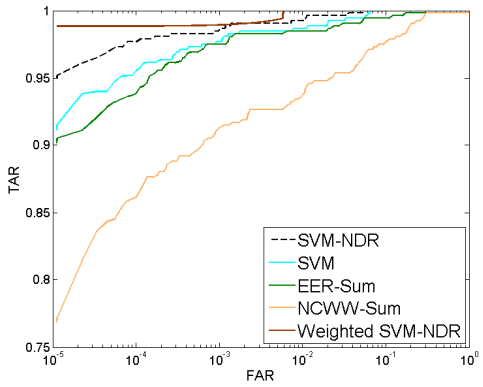
5.3.5 Results

The performance achieved by the proposed solution and the baseline approaches is presented as EER values in Table 5.1 and as ROC curves in Figure 5.4. Performance was presented for all possible bi-modal combinations as well as for the fusion of all available sources (two face and two fingerprint matchers).

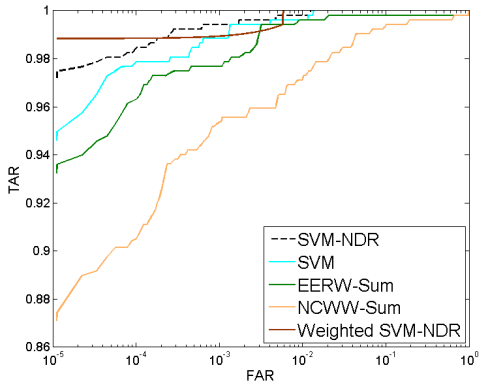
	Weighted SVM-NDR	SVM-NDR	SVM	EERW-Sum	NCWW-Sum	Face	Finger
Fc-Fli	**0.00552**	0.00582	0.01133	0.01338	0.04069	0.04550	0.08669
Fc-Flr	0.00581	**0.00397**	0.00406	0.00491	0.02015	0.04550	0.04674
Fg-Fli	**0.00531**	0.01148	0.01156	0.02642	0.01194	0.05801	0.08669
Fg-Flr	0.00523	0.00976	**0.00422**	0.01404	0.00425	0.05801	0.04674
All	**0.000067**	0.00012	0.00167	0.00222	0.00371	X	X

Table 5.1: the EER values achieved by the proposed solution and the discussed baseline approaches on the different bi-modal biometric combinations, as well as the fusion of all four available sources. In the last two columns are the EER values of the single biometric sources involved in the fusion.

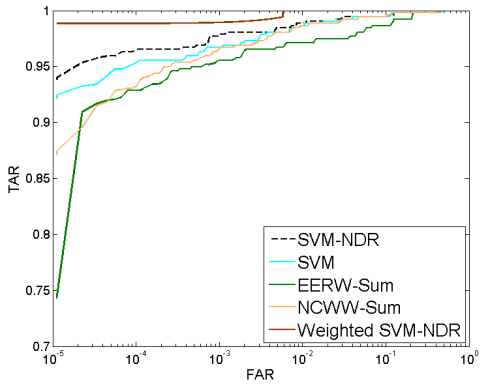
The EER values stated in Table 5.1 show the positive influence of integrating the NDR values (especially the weighted ones) in the fusion process. The proposed solutions significantly outperformed the baseline approaches under most of the experiment settings. The influence of the NDR integration in the fusion process is especially clear when dealing with the fusion of all four biometric sources. However, the main advantage of the proposed solution is providing high true acceptance rate (TAR) at very low false acceptance rate (FAR) values (below EER). This is clear in Figure 5.4e where all the ROC curves produced by different combinations of modalities show the high performance of the proposed solution at very low FAR values. The effect of the weighted integration of NDR values is also clear in the ROC curves by improving the performance at low FAR values, which are crucial for security related applications.



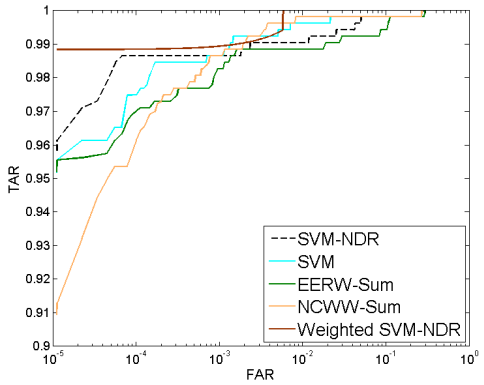
(a) Fc and Fli



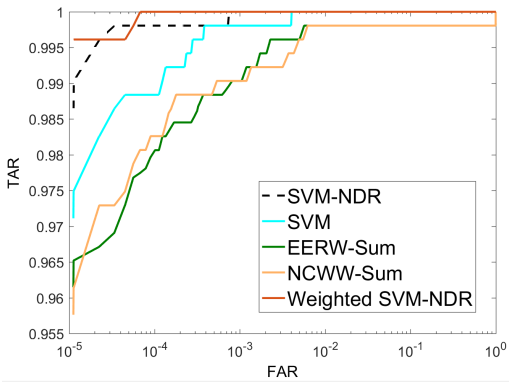
(b) Fg and Fli



(c) Fc and Flr



(d) Fg and Flr



(e) Fusion of all four sources (Fc, Fg, Fli, and Flr)

Figure 5.4: ROC curves achieved on the BSSR1 database: the curves shown here are for all possible bi-modal combinations of face matchers (Fc and Fg) and finger matchers (Fli - Flr) in the BSSR1 database and the results achieved by the fusion of all four available sources.

5.4 Multi-biometric score coherence

This section proposes embedding score coherence information in the fusion process to further enhance the multi-biometric performance. This follows the assumption that a minority of biometric sources, pointing out a different decision than the majority, might have faulty conclusions and should be given relatively smaller role in the fusion process. The formulation and evaluation of the proposed approach are presented in the following.

5.4.1 Defining score coherence

The score-level multi-biometric fusion approach presented in this work aims at integrating supplementary information based on the coherence of the fused scores. The coherence here points out the level of agreement of one score with all the other fused scores. The basic assumption is that, in a group of decision makers (multi-biometric sources) giving an opinion (score) on a certain topic (multi-biometric comparison), the mainstream opinion (opinion pointed out by the majority of decision makers) has a higher probability of being correct. Odd (outlier) decision, made by a relatively small number of decision makers, has a higher probability of being misinformed or misanalysed decision (e.g. noisy capture, poor preprocessing).

Within this scheme, a biometric score that has a higher level of agreement (coherence) with the other scores, in the same multi-biometric comparison, will be appointed a relatively higher weight and thus has more influence on the final decision. This is coupled with a static weight that points out the general quality of each biometric source.

The coherence measure for a certain score should point out the agreement of this score with all other scores in the same multi-biometric comparison. Based on this, a simple coherence measure was defined as the inverse of the average distance of the concerned score to all other scores in the multi-biometric comparison. Given that all scores are properly normalized, the coherence measure of the score $S_{k,l}$ (belonging to the source k out of K sources) in a multi-biometric comparison noted by l, is given as

$$Coh(S_{k,l}) = \frac{K-1}{\varepsilon + \sum_{i \neq k} |S_{k,l} - S_{i,l}|}, \tag{5.5}$$

where ε is a small positive number to avoid zero denominator (here, $\varepsilon = 0.01$).

A score with a higher coherence value points out a higher probability for a score to be of the mainstream decision of the multi-biometric sources, and thus, should be given a relatively higher weight. This results in the coherence based dynamic weight given as

$$w_{k,l}(Coh_{k,l}) = \frac{Coh_{k,l}}{\sum_{i=1}^{i=K} Coh_{i,l}}. \tag{5.6}$$

5.4.2 Static weights

To influence the general accuracy of each biometric source in the multi-biometric fusion process, static weights are used to weight the biometric scores. The static weights are constant for each biometric source (hence, static). In this work, the static weighting is used as a baseline solution to measure the effect of adding the coherence information into the fusion process. They are also used to influence this information along with the coherence based dynamic weights as will be shown later. Three different types of static weights w_k are used, namely the equal error rate weighting (EERW), the D-Prime weighting (DPW), and the Fisher discriminant ratio weighting (FDRW). These weighting approaches are discussed in more details in Section 4.2.

5.4.3 Fusion

To capture both the general performance of each biometric source and the individual certainty represented by the coherence, a combined weight was proposed as follows

$$w_{k,l}(Coh_{k,l}, St_k) = \beta w_{k,l}(Coh_{k,l}) + (1-\beta) w_k(St_k). \tag{5.7}$$

Here, β is a constant between zero and one [0,1]. β controls the relative effect of the dynamic and static weights. Different values of β are evaluated to optimize the tradeoff between both types of weights. *St* is the static weight parameter that can be EER, DPW, or FDR.

The fused score based on the dynamic weighting is given by

$$F = \sum_{i=1}^{K} w_{i,l}(Coh_{i,l}, St_i) S_{i,l}, \tag{5.8}$$

where $S_{k,l}$ is a score of the biometric source k of the comparison l and $w_{k,l}$ is its corresponding dynamic weight as in Equation 5.7.

5.4.4 Experimental setup

Database: the database used to develop and evaluate the proposed solution is the BioSecure multi-modal biometric database [OFA*10]. This database was acquired within the framework of the European BioSecure Network of Excellence. This work utilizes three biometric sources out of the DS2 part of the BioSecure database, the face (webcam, no flash) and both the left and right middle fingers captured by an optical sensor. This data was collected on a desktop PC environment in seven different European institutions and totalled in 210 subjects over two sessions.

Noise simulation: to simulate a more realistic scenario, the raw captures of both fingers and face images were subjected to blurring using an averaging filter of the size *m x m*. The blurring was performed on the second session data, considered as probe in this work. While the data of session 1 was considered as reference data and was not subjected to additional noise. The noise was applied by randomly selecting the filter dimension m to be one of $\{7, 9, 11, 13\}$.

Each probe sample of the three biometric sources was compared to each reference sample resulting in a similarity score. This was done using the original (noise-free) data and the data with induced noise. This resulted in a noise-free and a noise-induced scores databases. To create a realistic scenario, a certain percentage of the noise-free score database was randomly replaced by scores from the noise-induced score database. This resulted in the four score databases used in this work with 0%, 2.5%, 7.5%, and 15% of their scores originating from the noise-induced probes.

Biometric comparators: the following methods were used to measure the comparison score between pairs of face images and pairs of fingerprint scans:

Face comparison: to calculate a similarity scores between face captures, the OpenFace implementation was used [ALS16]. OpenFace is a Python and Torch implementation based on the work of Schroff et al. [SKP15]. This solution utilizes a deep neural network to build a 128-dimensional unit hypersphere face representing.

Fingerprint comparison: fingerprint comparison used the NIST Biometric Image Software (NBIS) implementation [WGT*07]. This implementation utilized the MINDTCT algorithm [WGT*07] to locate all minutiae in a fingerprint scan, assigning to each minutia point its location, orientation, type, and quality. BOZORTH3 algorithm [WGT*07] is used to perform the comparison by using the minutiae detected by MINDTCT to determine if two fingerprints are from the same person and same finger.

Experiments: the goal of the experiments is to show the effect of embedding coherence information in the multi-biometric score-level fusion process. This effect is also important in the more realistic scenario where some captured data is noisy. To achieve that, the proposed solution and the baseline solutions are tested on four different database settings. As described in Section 5.4.4, the databases always included noise-free reference data and a certain percentage of noise-induced probe data. The four resulting score databases contain a portion of scores originated from noisy probe data of the percentage 0%, 2.5%, 7.5%, and 15%.

To evaluate the statistical performance of the proposed solutions, the database was split into three equal-sized partitions. Experiments were performed on all possible fold combinations where one partition is used as an evaluation set and the other two are used as a development set. All the reported results are the averaged results of the three evaluation/development combinations.

Min-max normalization was used to bring comparison scores produced by different biometric sources to a comparable range. Min-max normalized score is given as in Equation 4.12. The evaluation was performed on each of the four score databases using the three different static weights EERW, FDRW, and DPW. As in Equation 5.7, β was used to control the relative effect of each of the dynamic coherence weight and the static weight. A β value of zero presented the baseline solution where only the static weight is in effect. The fusion included the three biometric sources (face and two fingerprints) under a verification scenario.

5.4.5 Results

The achieved results under different experiment settings are presented as ROC curves and EER values. The EER is the common value of the FAR and false rejection rate (FRR) at the operational point (decision threshold) where both rates are equal. The EER value provides a general and comparable measure of the evaluation performance, lower EER values correspond to higher performance. ROC curves plot the FAR and the TAR at different operational points (thresholds) and present the tradeoff performance between the two rates. In contrast to EER values, ROC curves provide a wider insight into the verification performance at all possible operational points. This might be of interest for a user focused on a relatively low FAR or FRR rate for a specific application.

The EER values achieved under different experiment settings are presented in Table 5.2. The positive effect of including coherence information becomes apparent when comparing the baseline approaches ($\beta = 0$) at different noisy data percentages. When no noisy data was involved, EER was reduced by 48% when combined with EERW and 82% when combined with FDRW, both at $\beta = 0.85$ and in comparison with the baseline pure static weighting ($\beta = 0$). At a noisy data percentage of 15%, the EER reduction was 63% and 81% for the EERW and FDRW based solution respectively (comparison between $\beta = 0$ and $\beta = 0.85$). It must be noticed that although the effect of the noise is clear in the baseline static weight solutions, the achieved error rates when including coherence information are only slightly affected by noise.

To put the presented values in perspective, the single source EER value for the face source is 2.0% in the 0% noise-induced data and 4.3% in the 15% noise-induced data. The EER values for the left middle fingerprint source are 3.3% and 4.6% respectively.

The ROC curves in Figure 5.5 show the effect of including the coherence information at different operation points. The curves are a comparison of the EERW based solution at the baseline $\beta = 0$ and the presented

coherence based solution at $\beta = 0.85$. The improvement in the performance at very low FAR is clear on noise-free data. More importantly, the performance of the coherence based solution on the noise-induced data almost matches that of the noise-free data. On the other hand, the negative effect of the more realistic noise-induced data on the performance of the baseline solutions is noticeable. Similar behaviors also appear for the FDRW and DPW based solutions.

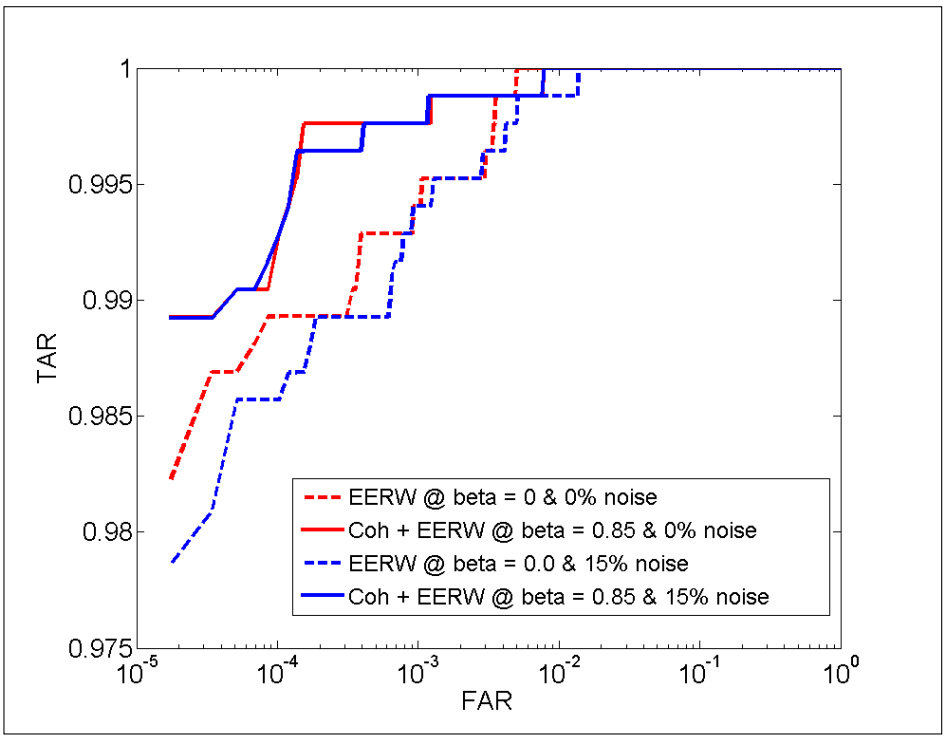


Figure 5.5: ROC curves showing the performance of the proposed solution and the baseline using the EERW as static weight under different percentages of noisy data.

	EERW Percentage of noisy data				DPW Percentage of noisy data				FDRW Percentage of noisy data			
β	0.00%	2.50%	7.50%	15.00%	0.00%	2.50%	7.50%	15.00%	0.00%	2.50%	7.50%	15.00%
0	0.2349	0.2349	0.3546	0.3454	0.2535	0.253	0.3709	0.358	0.6691	0.6694	0.795	0.7197
0.05	0.2307	0.2307	0.3491	0.3391	0.2473	0.247	0.3643	0.3574	0.5861	0.5861	0.6971	0.6837
0.1	0.2218	0.2218	0.3414	0.3343	0.2381	0.2375	0.3589	0.3543	0.4971	0.4968	0.4991	0.4087
0.15	0.2143	0.2141	0.3371	0.3331	0.2318	0.2315	0.3506	0.348	0.4553	0.4544	0.4851	0.384
0.2	0.2103	0.2098	0.33	0.2684	0.1686	0.1683	0.2865	0.3431	0.3703	0.3697	0.4728	0.38
0.25	0.1439	0.1434	0.2639	0.2596	0.1625	0.1623	0.2816	0.334	0.2776	0.277	0.3903	0.3643
0.3	0.1402	0.1402	0.2618	0.2621	0.1514	0.1508	0.2681	0.332	0.2598	0.2596	0.378	0.3609
0.35	0.1365	0.1362	0.2613	0.2581	0.1437	0.1437	0.2699	0.2664	0.2435	0.243	0.3663	0.3557
0.4	0.1316	0.1316	0.259	0.2553	0.1425	0.1434	0.2598	0.2604	0.2398	0.2361	0.3586	0.3514
0.45	0.1328	0.1331	0.2455	0.2418	0.1394	0.1394	0.2581	0.2547	0.2318	0.2315	0.3494	0.3437
0.5	0.1331	0.1319	0.2338	0.2309	0.1334	0.1331	0.243	0.2395	0.2195	0.2192	0.338	0.3348
0.55	0.1265	0.1299	0.2238	0.2232	0.1322	0.1311	0.2292	0.2264	0.2066	0.2109	0.3288	0.271
0.6	0.1282	0.1273	0.1574	0.1551	0.1305	0.1294	0.2206	0.2189	0.1422	0.1431	0.2533	0.255
0.65	0.1276	0.1268	0.15	0.1465	0.1276	0.1268	0.1505	0.1477	0.1328	0.1328	0.2307	0.2378
0.7	**0.1219**	**0.1213**	0.1411	0.1402	**0.1236**	0.1276	0.1457	0.1439	0.1302	0.1296	0.2189	0.2204
0.75	**0.1242**	**0.1236**	0.1336	0.1334	**0.1228**	**0.1222**	0.1334	0.1379	**0.1245**	0.1305	0.1522	0.1497
0.8	0.1262	0.1253	0.1322	0.1314	**0.1248**	**0.1239**	**0.1311**	0.1296	**0.1251**	**0.1242**	0.1431	0.1405
0.85	**0.1219**	**0.1211**	**0.1259**	**0.1294**	**0.1216**	**0.1208**	**0.1314**	**0.1294**	**0.1216**	**0.1208**	**0.1348**	**0.1359**
0.9	**0.1251**	**0.1242**	**0.1268**	**0.1245**	**0.1248**	**0.1239**	**0.1265**	**0.1253**	**0.1219**	**0.1211**	**0.1256**	**0.1296**
0.95	0.1262	0.1256	**0.1294**	**0.1253**	0.1259	0.1253	**0.1291**	**0.1248**	0.1256	**0.1251**	**0.1282**	**0.1245**
1	0.1268	0.1265	**0.1314**	**0.1242**	0.1268	0.1265	0.1314	**0.1242**	0.1268	0.1265	**0.1314**	**0.1242**

Table 5.2: the achieved EER values (in percentage) for the different baseline solutions and with different levels of the proposed score coherence influence (β) under different percentage of noise-induced data. The lowest range of error rates per experiment setting (column) are in bold.

5.5 Summary

This chapter aimed at improving the performance of multi-biometric systems by proposing including supplementary information in the multi-biometric process. The proposed solutions aimed at being independent of the raw captured data or having prior knowledge of the capture environment/device. Two types of supplementary information were proposed, integrated, and evaluated.

Score-level multi-biometric fusion solution conventionally depended solely on the comparison scores from multiple sources. Advanced solutions considered information about the quality of the captures and/or variations in the capture environment. Including these types of supplementary information had a positive effect on the multi-biometric performance. However, such information requires either prior knowledge (e.g. of capture environment) or consistent and performance related quality estimators.

First, this chapter proposed the introduction and integration of neighbors distance ratio into the fusion process. This is based on the assumption that a genuine score of a certain probe subject is relatively distanced from the set of a clustered set of imposter scores produced by the same probe. As a result, the genuine/imposter decision did not only depend on the comparison scores from multi-biometric sources, but also on the relation to other comparisons within a 1:N biometric comparison. This integration was implemented in a classification-based fusion approach that utilized support vector machines. An improved solution was also presented by integrating biometric source weighting information in the NDR calculation process. The evaluation was performed on the BSSR1 database and the achieved results showed that integrating NDR information can largely improve the multi-biometric fusion performance, especially at the critical low FAR range. Different experiments presented an NDR induced reduction of the FRR by more than 50% at the law FAR value of 0.01%. These observations encourage the consideration of the proposed NDR information in large-scale, performance critical applications such as duplicate enrollment checks and black list identification.

The second type of supplementary information proposed in this chapter is the comparison score coherence. This was based on the assumption that the minority of decision makers (biometric sources) pointing out a different decision than the majority, might have faulty conclusions and should be given a relatively smaller role in the final fused decision. This was incorporated in a dynamic weighting approach that also considers static weights. The approach was evaluated on a database with different levels of induced noise and was compared to three baseline static weight solutions. Including the coherence information proved to largely enhance the biometric performance, especially in the more realistic scenario where some of the captured data could be slightly noisy. EER values under different test scenarios were reduced by at least 45% as a result of introducing coherence information.

This chapter responded to *RQ5* by proposing a measure derived from the relations between different comparisons and proving that this information can be used to discriminate between imposter and genuine comparisons and can improve the biometric performance. It also responded to *RQ6* by proposing a coherence measure between multi-biometric sources and integrating it into the fusion process, which proved to increase the final decision accuracy. The next chapter will tackle a different component in the multi-biometric work-flow, the reference database and its management.

6 Multi-biometric data retrieval

The last two chapters (4 and 5) were concerned with the optimization of the fusion process by focusing on multi-biometric source weighting and introducing supplementary information into the fusion process. This chapter discusses a different component of the multi-biometric work-flow, the reference database. Indexing of multi-biometric data is required to facilitate fast searchs in large-scale biometric systems. Previous works addressing this issue were challenged by including biometric sources of different nature, utilizing the knowledge about the biometric comparisons, and optimizing and tuning the retrieval performance. This chapter presents a generalized multi-biometric retrieval approach that adapts the Borda count algorithm within an optimizable structure. This chapter is based on the published works [DTBK17a, DTBK17b, DTBK17c].

6.1 Introduction

Large-scale biometric systems are spreading to facilitate security and service goals worldwide. An example of such a system is the e-Aadhaar project of the unique identification authority of India (UIDAI) [e-A15], with the goal of enrolling 1.2 billion citizens. Building such a system requires trillions of biometric comparisons within duplicate enrollment checks to insure single enrollment per person. Duplicate checks and identification queries in such a system are challenged by the limitation of computational efficiency as they require exhaustive comparisons.

Indexing techniques aim at reducing the number of comparisons (candidate identities) required by an identification system with a large number of biometric references. The fuzziness of biometric data makes the indexing task quite challenging. The availability of multi-biometric records in large-scale biometric systems offers the chance to produce more accurate indexing approaches that lead to faster searches.

Iris biometrics is a realistic multi-biometric use case as the capture process usually samples both irises at once. Iris texture is rich in information, including spots, rifts, colors, filaments, and minutia. It has about 10^{72} possible patterns [PA07] that makes it very unique and generally result in one of the smallest false-matching rates of all biometric traits [APM*12]. These properties make the iris an outstanding candidate for large-scale-biometric systems and an example use case for the multi-biometric indexing solution presented in this work.

Previous multi-biometric indexing solutions focused either on the feature or the rank-level fusion. The feature-level poses limitations on the possible combinations of modalities and algorithms used by each biometric source. Moreover, it might face challenges in situations where some sources are missing. Solutions focusing on the rank-level are more flexible in terms of different biometric sources and missing data. However, these solutions disregard in-depth information and treat every rank and every biometric source equally. Using Borda count to create a multi-biometric indexing solution uses more of the available information (ranks), but assumes a linear importance (weight) loss as it goes down the ranks and is not optimizable for different biometric sources. Moreover, none of the previously discussed solutions offered the possibility to focus the performance optimization on a certain operational range.

This chapter proposes a generalized solution based on the Borda count that can be optimized for different biometric sources. Moreover, it can be tuned to achieve better results at certain operating points. The proposed

solution is also extended to use approximate index distances to create a more accurate retrieval, even when dealing with low quality candidate lists form single biometric sources. A comparison is made with baseline approaches in terms of general indexing performance, required query time, performance on mixed quality candidate lists, and different levels of missing data.

In Section 6.2, a detailed look into related works is presented. Section 6.3 discusses the baseline and the proposed solutions including its theoretical background. The experiment setup and the achieved results are detailed in Sections 6.4 and 6.5. A final discussion is drawn in Section 6.6.

6.2 Related work

Different feature extraction approaches were proposed for biometric iris representation. However, some of the most accurate and widely-used approaches, such as the Daugman iris codes [Dau04] and the ordinal measures (OM) [ST09], suffer from rotational-inconsistency inherited from the sensitiveness to eye tilt. This has limited the possibilities of developing accurate and fast indexing structures for iris databases. Recently, a number of rotation-invariant feature transformations were proposed with an aim to enable iris indexing [DTBK17a, RBBB14].

An indexing structure is a data structure that is used to quickly locate where an index value occurs. To perform fast identification, only the neighborhood of the query index has to be searched to reduce the search space drastically. Unfortunately, biometric data has no natural order by which one can sort it and thus indexing biometric data is a challenging task.

Driven by the demand for large-scale biometric systems, different approaches were proposed to reduce the response time for iris identification. Daugman *et al*. [HDZ08] proposed a fast search algorithm based on Beacon Guided Search on iris codes using the multiple colliding segment principle, which results in low query times but needs complex memory management. Mehrotra *et al*. [MSMG09] proposed an indexing algorithm that divides the iris image into subbands, then create a histogram of transform coefficients for each subband. A key is created based on these histograms, then organized into a search tree achieving a hit rate 98.5% at a penetration rate of 41%. Mukherjee and Ross [MR08] proposed two indexing techniques for both iris codes and iris textures, which achieve hit rate of 84% at 30% of the search space. One of the most accurate tree based approaches was presented by Jayaraman *et al*. [JPDG08]. By using principal component analysis (PCA) in combination with B+ trees, a hit rate of 93.2% was achieved along with a penetration rate of 66.3%.

Gadde *et al*. proposed a technique based on Burrows Wheeler transformation reaching a hit rate of 99.8% while reducing the search space to 12.3% [GAR10] on an evaluation database containing only 249 subjects. In an indexing method proposed by Rathgeb and Uhl [RU10], the search space could be reduced to 3% while reaching a hit rate around 90%. This is achieved by generating 4-bit biometric keys from the iris image to use it as a starting position in a Karnaugh map. However, these results go along with very high storage cost. More recently, Rathgeb *et al*. proposed an iris indexing approach based on Bloom filters achieving a hit rate of 93.5% at 6.2% penetration rate [RBBB15]. However, this approach uses all samples at every tree level and it requires a full tree replacement for any deletion operation in the database, with a complexity of $\mathcal{O}(N\log(N))$. More recently, an iris indexing approach was proposed based on locality sensitive hashing forests (LSH-Forest) and rotation invariant iris representation [DTBK17c]. This achieved a hit rate of 99.7% at a 0.1% penetration rate with logarithmic complexity of query and storage requirements that grows linearly with the database size.

Fewer works addressed the promising aspect of multi-biometric indexing. Gyaourova and Ross proposed solutions for multi-biometric indexing of face and fingerprint biometrics [GR09, GR12]. They created index codes based on the comparison scores of an input with a small constant set of references. Three approaches were evaluated, namely the union and intersection of both candidate lists, as well as the concatenation of both index

codes. The union of the candidate lists performed the best with a hit rate of around 99,5% at a penetration rate of 5%, compare to around 92.5% and 90% hit rate for face and fingerprint respectively at the same penetration rate. However, this was achieved with the large index code of 256 dimensions. Gyaourova and Ross [GR09, GR12] also proposed the use of a number of rank-level multi-biometric indexing schemes that achieved worse results compared to index code concatenation, these schemes will be discussed as part of the baseline approaches in this work (Section 6.3.2). The concatenation of iris features was used to create a multi-instance iris indexing achieving 99.98% hit rate at a 0.1% penetration rate [DTBK17c]. However, the index concatenation approach is not extendable to multi-biometric systems using different modalities or indexing approaches, and is sensitive to missing data (not all sources are available).

6.3 Methodology

6.3.1 Single iris indexing

A solution based on OM features [ST09], rotation invariant representation (RIR) [DTBK17a], and locality sensitive hashing forest (LSH-Forest) based indexing [DTBK17c] was used as the basic single-iris indexing approach.

Ordinal measures represent iris images as a binary iris code. They describe a quality measurement related to the relative ordering of several quantities. Given two distinct image regions, the ordinal measure between these regions is encoded by the inequality of the average intensities. By applying this measure multiple times to different regions, a code can be generated. With multilobe differential filters (MLDFs) the OM features can additionally be extracted with flexible interlobe and intralobe parameters, such as location, scale (intra), orientation and distance (inter). To extract ordinal iris features using MLDFs, an MLDF operator slides across the whole normalized iris image and each ordinal comparison is encoded as one bit. By concatenating all binary bits of the image, an iris code is generated, which is also called ordinal code. The challenge still facing iris features, such as OM, is the rotation inconsistency of the resulting iris code. Ordinal measures feature extraction approach will be considered as a basis for the rotation-invariant representation presented in this work.

Rotation invariant representation: to build an accurate biometric indexing structure, indices have to be extracted from discriminant and compact representations of the biometric characteristics. Iris biometrics inherently suffers from rotation inconsistency, which affect indexing efforts. Based on this, our indexing scheme will be based on the recently proposed rotation invariant, accurate, and compact *RIR transformation* [DTBK17a]. This transformed representation will be based on the binary OM features extracted from iris images [ST09].

The RIR transformation consists of a linear combination of two basic transformations $\hat{u}(\mathbf{v})$ and $\hat{v}(\mathbf{v})$. Given a binary iris code $\mathbf{v} \in \{0,1\}^n$ of length n, the $\hat{u}_k$ basic transformation describes how many pairs of 0's have a distance of k within $\mathbf{v}$, while $\hat{v}_k$ specifies the same for pairs of 1's. Now, the RIR transformation is defined component-wise as

$$\mathrm{RIR}_k(\mathbf{v}) = \hat{u}_k + \hat{v}_k = \sum_{i<j} \delta(d(i,j) - k)\, \delta_{v_i, v_j}, \tag{6.1}$$

where δ is the Kronecker delta and the distance $d(i,j)$ between the i^{th} and j^{th} location of $\mathbf{v}$ is defined as

$$d(i,j) = \min\{|i-j|, n-|i-j|\}. \tag{6.2}$$

As a result, the k^{th} component of RIR(v) is the number of same labeled pairs with a distance of k within $\mathbf{v}$. This leads to a rotation invariant representation of the iris that enables our proposed indexing structure.

LSH-Forest is used to create indexes based on the RIR features. The main goal of locality sensitive hashing (LSH) is to perform a projection into a lower dimensional space where if similar high-dimensional vectors are projected, their projections to the lower dimensional subspace are similar too [SC08]. The LSH index suffers from the need of tuning, the continuous retuning and the lack of a quality guarantee for all queries. However, this can be solved by some adaption leading to *LSH-Forest* [BCG05]. LSH-Forest proved to produce an accurate and easily maintainable indexing structure, with the complexity of insertion and deletion is limited to the same logarithmic complexity of a query, and the required storage grows linearly with the database size [DTBK17c].

6.3.2 Multi-biometric indexing

given a biometric system consisting of m biometric sources (e.g. modalities), each with their own indexing structure. If a query q is forwarded to the system, each source (modality) j returns a ranked candidate list CL_j of length $|CL| = l$ consisting of the most probable corresponding identities in sorted order. The rank $r(i)$ describes the rank of identity i in the list, where $r(i) = 0$ means that i is at the first place whereas $r(i) = l$ refers to the case in which i is not in the list. In order to be independent of the size of the candidate list, a normalized rank $\hat{r} = r/l$ is used in this work.

Baseline indexing solutions: for baseline indexing solutions, several approaches based on rank-level fusion and feature fusion are used. At feature-level, *concatenation* describes the approach in which each feature vector of an identity is simply concatenated in a fixed order to create an extended feature vector. In this work, the same approach used as the basic single iris indexing, based on LSH-Forest, is used on the concatenated vector to produce the multi-biometric index. This approach limits the possibility of integrating different modalities with different comparator styles (e.g. fingerprint and iris) and may face challenges dealing with missing data (subset of query data).

At the rank-level, Gyaourova and Ross described two methods in which they treat the candidate lists as sets instead of rankings [GR09, GR12]. The first is the *intersection* fusion scheme that computes the intersection of the candidate lists of the individual sources as the final list $\mathcal{C} = \cap_j CL_j$ in order to reduce the size of the search space. However, a weak performance of a single source can result in the loss of the searched identity. Similar to that, the *union* fusion scheme outputs the union of the candidate lists as the final list $\mathcal{C} = \cup_j CL_j$ to prevent a weak performance of a single modality from affecting the overall performance. The *highest rank* method is another rank-level scheme where a final candidate list is produced by only considering the highest ranks of each identity allowing to focus on the strength of each source. For each identity i, the highest (minimum) rank $R(i) = \min\{r_1(i), \ldots, r_m(i)\}$ is computed and the final list of identities are outputted in ascending order of their highest ranks. Here, the major risk is the high number of occurring ties.

In the context of group decision theory, the *Borda count* method was introduced [HHS94] as a generalization of the majority vote. For each identity i, a Borda count $B(i) = \sum_j r_j(i)$ is computed to measure the magnitude of agreement of all sources that i corresponds to the searched identity. By outputting the identities in ascending order, the final list is produced. The major disadvantage is the assumption of the additive independence of the contributing sources that does not take into account the differences of the individual indexer capacities and the performance of each of the sources.

In conclusion, this work considers five baseline solutions. Namely, the feature *concatenation*, rank *intersection*, rank *union*, *highest rank*, and the *Borda count*.

Proposed Generalized solution: the proposed solution is designed to work on the rank-level to enable a more liberal integration of different modalities and algorithms, which can have many restrictions at the feature-level. It also aims at enabling a generalized solution that can take more advantage of the available information about individual sources.

General Borda count (GBC): the proposed multi-biometric indexing solution aims at optimally including information about the performance of the separate biometric sources indexing approaches. The proposed indexing fusion method is based on a cumulative score function

$$cs(i\,|\,\vec{\mu},\vec{\varphi}) = \sum_{j=1}^{m} s\left(\hat{r}_j(i)\,|\,\mu_j\right) + \rho\left(\hat{r}_j(i)\,|\,\varphi_j\right), \tag{6.3}$$

which computes a value cs for the identity i based on its ranks in each source (modality) and consists of a non-linear weighting term $s(\hat{r}_j(i)\,|\,\mu_j)$ with weighting parameter μ_j and a penalty term $\rho(\hat{r}_j(i)\,|\,\varphi_j)$ with penalty coefficient φ_j. Thereby, $\hat{r}_j(i)$ returns the normalized rank of identity i given by source j. The non-linear weighting term is given by the score function

$$s(\hat{r}|\mu) = \begin{cases} s_c(\hat{r}|\mu) & \text{if } \mu \geq -1 \\ s_u(\hat{r}|\mu) & \text{if } \mu < -1 \end{cases}, \tag{6.4}$$

where

$$s_c(\hat{r}|\mu) = (1-\hat{r}) \cdot e^{-\mu\hat{r}}, \tag{6.5}$$

describes the common score function for indexing structures having a high probability that the corresponding identity is within the first ranks and

$$s_u(\hat{r}|\mu) = -\mu(1-\hat{r}) \cdot e^{\mu(1-x)+1}, \tag{6.6}$$

is the uncommon score function for indexing techniques with a weaker performance. The penalty term

$$\rho(\hat{r}|\varphi) = -\varphi \cdot \delta(\hat{r}-1), \tag{6.7}$$

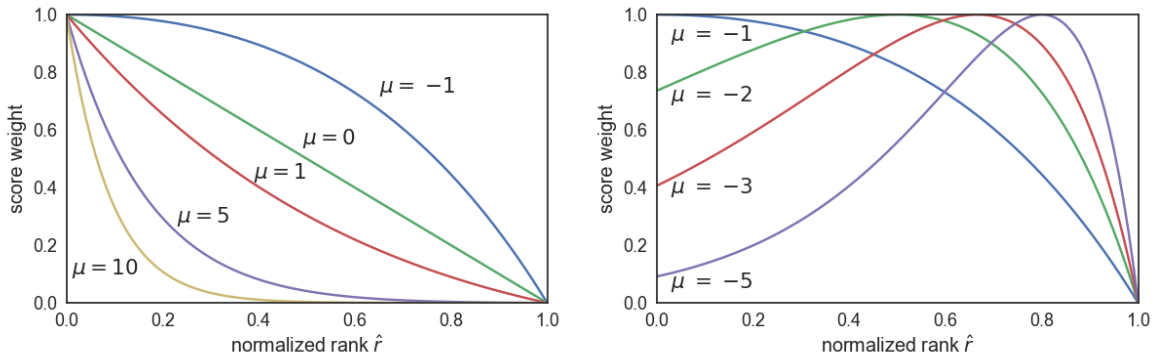
returns a negative value $\varphi \in [0, m-1]$ if the identity is not in the list, and therefore penalizes candidates outside of candidate lists. As a result, identities frequently occurring in lists are getting higher scores and will be at lower ranks in the final candidate list, which is particularly useful when sources perform weakly.

During retrieval, each identity i in the lists is assigned to a score $cs(i\,|\,\vec{\mu},\vec{\varphi})$ based on the source parameters and its ranks in the different sources. By outputting the identities in decreasing order of their cumulative scores, a final candidate list is produced.

Demonstrating the behavior: in Figure 6.1a, the behavior of the score function $s_c(\hat{r}|\mu)$ for $\mu \geq -1$ can be seen for a single source. At high values of μ, candidates with lower ranks get high weights compared to the rest. As a result, these candidates get preferred in the fused list and the behavior of the highest rank method is imitated. By varying parameter μ, this preference of the lower ranks over the higher ones can be controlled. The special case $\mu = 0$ refers to a linear weighting and results in the original Borda count method. For $\mu \in [-1, 0)$ the descent of candidates weights is lower at the beginning, which leads to weak preferences between the first candidates.

In Figure 6.1b, the behavior of the score function $s_u(\hat{r}|\mu)$ for $\mu < -1$ can be seen. This function is particularly useful if the indexing structure has the highest probability to return the corresponding identity at rank $\hat{r} \neq 0$. Therefore, the score function prefers candidates around rank $\hat{r} = 1 + \frac{1}{\mu}$, while candidates outside the list ($\hat{r} = 1$) remain at a score of zero.

The penalty term $\rho(\hat{r}|\varphi)$ can be interpreted as a "soft intersection". While the original intersection method harshly penalizes a weak performing indexing structure, the penalization can be continuously controlled over the coefficient φ. Therefore, this approach can deal with weak performing indexing structures and allow eliminating candidates that are not frequently considered by different sources.



(a) (b)

Figure 6.1: score functions behavior: a) Score function $s_c(\hat{r}|\mu)$ for different values of $\mu \geq -1$. b) Score function $s_u(\hat{r}|\mu)$ for different values of $\mu \leq -1$.

General distance Borda count (GDBC): this approach aims at utilizing more in-depth information compared to ranks. This information is the efficiently calculated approximate distances between the query and references indexes. Many indexing structures allow returning approximate distances to the query along with a candidate list. These approximate distances can be calculated using their multi-dimensional indexes only. Therefore, they offer a way to gather more information without increasing the computation time significantly. In order to use these distances d, they need to be normalized. This is done via min-max normalization of a candidate list CL.

$$\hat{d} = \frac{d - \min\{CL\}}{\max\{CL\} - \min\{CL\}}. \tag{6.8}$$

Given biometric system with m indexing structures enable to produce approximate distances jointly with a candidate list. For each identity in the lists, a score

$$\overline{cs}(i\,|\vec{\mu},\vec{\varphi}) = \sum_{j=1}^{m} s\left(\hat{d}_j(i)\,|\mu_j\right) + \rho\left(\hat{d}_j(i)\,|\varphi_j\right), \tag{6.9}$$

is assigned, where $\hat{d}_j(i)$ refers to the normalized approximate distance of identity i given by source j. By outputting the identities in decreasing order of their cumulative scores, a final candidate list is produced.

The approximate distance used in this work takes advantage of the LSH that project vectors from a large (n-dimensional) space to a lower (h-dimensional) subspace using l hash functions. In the case of two binary vectors v and w, the distance between them can be calculated via hamming distance $HD(v,w)$. Under the condition $HD(v,w) \leq n-h$, which is satisfied for moderately large values of n, this distance can be estimated by

$$HD(v,w) \approx (n-h)\left(1 - \left(\frac{C_F(v,w)}{l}\right)^{\frac{1}{h}}\right),$$

where $C_F(v,w)$ counts the number of colliding hash functions [MS08].

Parameter optimization: in order to find the best operating point for a certain application, or in this case, to find the optimal parameters $\vec{\mu}$ and $\vec{\varphi}$, a quality measurement based on penetration and hit rate is needed. This

will allow the system developer to influence the indexing performance by focusing on achieving performance superiority at low penetration rates or at high hit rates with less regard to low penetration rate. Therefore, the cost for hit rate C_h and the penetration rate C_p are introduced, whose ratio, called cost ratio, is given by $\frac{C_h}{C_p}$. A large cost ratio refers to a scenario in which a low penetration rate is important, while a small cost ratio points to a high hit rate, usually achieved at higher penetration rates. For a given cost ratio $\frac{C_h}{C_p}$, the linear cost metric

$$LCM_{C_p}^{C_h}(c) = \max\left\{h - \frac{C_h}{C_p}p \,\middle|\, (p,h) \in c\right\}, \tag{6.10}$$

describes a evaluation score for a penetration-hit rate curve c, where p and h describe the corresponding penetration and hit rates of the curve analogous to the linear cost metric known from ROC curves.

By maximizing the linear cost metric $LCM_{C_p}^{C_h}$, the optimal parameters $\vec{\mu}$ and $\vec{\phi}$ for the general Borda count can be calculated for a given cost ratio. It is also possible to optimize the hit rate for a given penetration rate. However, if the size of the database is changing, the optimization process has to be done for a different penetration rate while the cost ratio is still the same for the same application.

6.4 Experimental setup

Database: the ISYN1 iris synthetic images database [CFM10, CFMT10, CMM02, ST09, WTS08] was used in this chapter to develop and evaluate the proposed solution. This database is generated by CASIA [Chi13] using its synthetic generator software [WTS08]. In this chapter, 10,000 reference and 10,000 probe iris images for each of the left and right irises were used.

Creating single iris index: the OM codes were extracted from each image in the database. These codes are transformed into a compact and rotation invariant space using the RIR transformation [DTBK17a] then each element in the resulting vectors is normalized using z-score normalization.

For single iris indexing, LSH-Forest was used, as described earlier, to build an indexing structure based on the normalized RIR codes. The LSH-Forest uses $l = 10$ trees and a maximum 32 labels (index size). The hash functions to create the labels are generated by using the random projection principle on cosine similarity measure.

Different experiment settings:

Indexing performance of different approaches: five baseline solutions discussed earlier were evaluated, feature concatenation, rank intersection, rank union, highest rank, and Borda count. The feature concatenation approach used the same settings as the single iris indexing with the number of tree used in the LSH-Forest $l = 10$ and a maximum of 32 labels. Additionally, to put the overall performance in perspective, the multi-biometric indexing approach presented by Gyaourova and Ross [GR09, GR12] was implemented and evaluated based on the RIR iris codes. The implementation used an index code dimensionality of 32 for direct comparison and the index code candidate list union, as the best performing multi-biometric approach presented in [GR09, GR12]. The proposed approaches, GBC and GDBC, were evaluated with different cost ratios (3-1, 1-1, and 1-3). The parameters ϕ and μ were chosen using simulated annealing by optimizing μ and ϕ for each source at the same time. The evaluation of the chosen parameters was performed via linear cost metric at a certain cost ratio.

Indexing performance vs. quality of candidate list: as the goal of this chapter is to propose a more generalized indexing solution, it is important to discuss the performance of different solutions on diverse accuracy cases. Based on this, a quality measure is defined to assign quality grades to retrieved (single biometric source) candidate lists. This will enable a discussion on the performance of different solutions on candidate lists of different quality grades.

grade name	grade	avg. rank ± SD	rank range
perfect	++	0.00 ± 0.00	0
good	+	6.07 ± 6.46	1 – 24
medium	0	36.00 ± 3.84	25 – 49
bad	–	74.82 ± 14.56	50 – 99
useless	– –	100.00 ± 0.00	100

Table 6.1: candidate list quality grade definition.

The quality grade is defined by the rank of the searched (correct) identity in a candidate list. This rank is grouped into five ranges to define five quality grades as in Table 6.1. Analyzing 1000 candidate lists (of the length 100) for each grade, the table shows the performance of single iris indexing for each grade by computing the average rank of the searched identity and its the standard deviation (SD).

The experiment analyzes how different rank-based multi-biometric indexing methods are able to handle candidate lists of different qualities. Therefore, source combinations are created by building pairs of all possible quality grades. Each time, the rank of the searched identity in the multi-biometric retrieved list is determined. This is repeated for every candidate list quality combination. Finally, the average rank and its standard deviation is computed.

Indexing performance vs. missing data: in scenarios where a query cannot provide information from all biometric sources, indexing solutions based on feature fusion methods show a susceptible behavior since they have to work around this situation. With regard to large-scale databases, missing data is a frequently occurring problem and indexing with rank-level fusion offers a simple and effective way to handle this issue. Therefore, it is interesting to analyze the performance of different multi-biometric indexing approaches on different level of missing data.

To simulate the missing data scenario, a percentage of reference templates was neglected randomly in a way that guarantees the existence of at least one valid single-source query per identity. As the experiment is performed on two multi-biometric sources, 50% missing data leads to having one biometric source only per each query. In the feature concatenation approach, this was dealt with by padding the feature vector by the average feature value (zero). The rank based approaches are inherently flexible to the number of sources and thus can deal with only one source. This was evaluated on the feature concatenation approach and compared to the proposed GBC and GDBC approaches.

6.5 Results

	Concatenation	Intersection	Union	HR	BC	GBC	GDBC	Single iris
Query time (ms)	29.7	37.1	36.5	35.1	36.1	36.6	37.2	16.8
HR (%)	99.72	98.80	99.73	99.73	99.70	99.77	99.78	98.16

Table 6.2: Query time and hit rate at 0.1% penetration rate of a 10 thousands references database

Indexing performance comparison: the achieved performance is presented as a relation between the penetration rate and hit rate. Here, the hit rate is the portion of the searches where the correct identity is found within the considered percentage of the references out of the complete database (penetration rate). Small penetration rates result in smaller number of in-depth comparisons required to make the final identification decision. The achieved

quality grade combination		Intersection			Union			Highest Rank			Borda Count		
++	++	0.0	±	0.0	0.0	±	0.0	0.0	±	0.0	0.0	±	0.0
++	+	1.4	±	3.0	0.0	±	0.0	0.0	±	0.0	0.2	±	0.8
++	0	9.4	±	11.1	0.0	±	0.0	0.0	±	0.0	2.5	±	4.2
++	−	22.9	±	27.3	0.0	±	0.0	0.0	±	0.0	8.2	±	10.7
++	− −	100.0	±	0.0	0.0	±	0.0	0.0	±	0.0	12.4	±	14.2
+	+	2.7	±	4.5	5.5	±	6.0	5.4	±	6.2	1.3	±	2.6
+	0	18.1	±	13.5	11.0	±	11.6	11.3	±	11.7	7.3	±	7.4
+	−	37.7	±	27.6	10.8	±	11.5	11.1	±	11.7	19.3	±	15.5
+	− −	100.0	±	0.0	11.0	±	11.6	11.4	±	11.8	31.4	±	17.4
0	0	21.9	±	13.0	54.6	±	10.7	54.5	±	10.5	19.7	±	13.2
0	−	43.7	±	25.5	58.3	±	10.3	58.7	±	10.3	43.0	±	17.6
0	− −	100.0	±	0.0	59.6	±	10.6	60.2	±	10.7	69.4	±	6.7
−	−	50.1	±	27.8	89.2	±	11.7	89.3	±	11.5	80.0	±	16.5
−	− −	100.0	±	0.0	94.4	±	9.4	94.6	±	9.2	96.2	±	6.6
− −	− −	100.0	±	0.0	100.0	±	0.0	100.0	±	0.0	100.0	±	0.0

Table 6.3: multi-biometric indexing performance given by the rank (average ± standard deviation) of the correct identity in the retrieved list for different candidate lists (single source) quality combinations - the baseline solutions.

performance by the five baseline solutions and the union of index codes approach are presented in Figure 6.2 with a comparison to the proposed solutions with the cost ration of 3-1.

Figure 6.3 illustrates the effect of tuning the cost ratio. It is noticed that when the cost ratio focuses on performance at lower penetration rate (cost ratio 3-1) the indexing performance is superior at low penetration values. On the other hand, when the cost ratio focuses on hit rate with low importance for the penetration rate (cost ratio 1-3), the proposed solution scores superior hit rates at higher penetration rates.

Query time: beside the tradeoff between the penetration and hit rates, the computational time required for a query is presented. The achieved query time is discussed to put the efficiency of the proposed approach in perspective as this implementation does not utilize the high possibility of computational parallelization and is measured on a desktop PC running on an Intel®Core™i5-4590 3.30 GHz CPU. Table 6.2 presents the query time required by the different approaches to retrieve a final candidate list at a fixed penetration rate (0.1%) of a database with 10,000 references, the achieved hit rate is also listed. It can be noticed that the proposed GBC and GDBC indexing do not require significantly higher query time compared to the baseline approaches while being more accurate. Feature concatenation requires slightly shorter query time compared to the proposed GBC approach, but with a significant loss in accuracy. GDBC requires slightly higher query time compared to GBC due to the time required for the approximate distance calculations.

Missing data: in realistic suboptimal scenarios, and due to various reasons, some of the single biometric sources may not be able to produce feature vectors. In such scenarios, the performance of a multi-biometric indexing solution has to be maintained. Therefore, missing data scenario was simulated by randomly neglecting a certain percentage of the biometric sources while maintaining a minimum of one source per comparison. As we are evaluating the use of two biometric sources, this means that a 50% missing data results in a situation where each query uses only one biometric source (iris). As the rank-based solutions inherently can deal with a missing candidate list, the plots presented in Figure 6.4 compares the performance of the feature concatenation approach

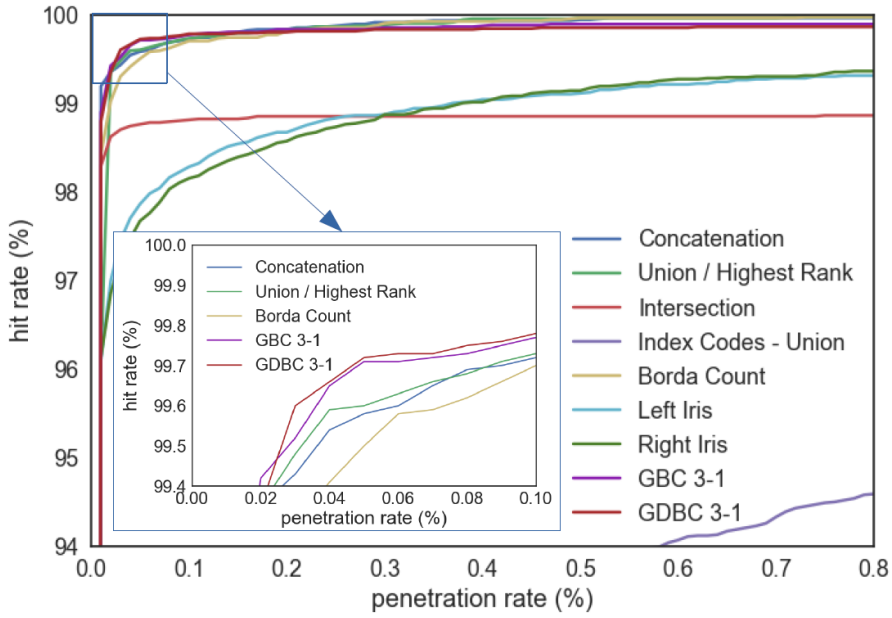


Figure 6.2: indexing performance of the proposed solutions compared to baseline solutions.

with the proposed GBC and GDBC approaches at different levels of missing data. It is noticed that the GBC and GDBC at 10% missing data over perform the feature concatenation even at 2.5% missing data at low penetration rates. The GBC and GDBC maintains the single iris indexing performance given 50% missing data, while the feature concatenation lags behind.

Quality of candidate lists: as described in Section 5.4.4, it is important to analyze indexing performance when presented by different candidate lists quality combinations. This gives a better idea on the indexing performance when used with variable biometric indexing sources and when dealing with less than optimal situations where the capture or pre-processing of the biometric data is of a variable quality. Here, we compare the baseline rank-based approaches with the proposed GBC and GDBC approaches by presenting the average position (and standard deviation) of the correct identity in a retrieved list of 100 identities. These values are presented for all possible combination of left and right iris candidate list quality grades. A lower average rank indicates a better multi-biometric indexing performance. Table 6.3 shows the achieved average ranks by the baseline solutions. As expected, single source candidate lists of low quality produce less accurate multi-biometric candidate lists.

Tables 6.4 and 6.5 presents the same information as in Table 6.3 but for the proposed GBC and GDBC respectively, and for different optimization variables ϕ and μ. Here, one can notice the high advantage of the GDBC approach that achieves superior indexing performance given low quality single source candidate lists as input. For example, given two single iris indexing candidate lists of the quality grades "$-$" and "$-$", the GDBC locates the correct identity in the multi-biometric index list in the rank 0.43 on average. Given the same lists, the GBC locates it in the 53.21 rank wile the intersection approach locates it in the 50.1 rank and the other baseline approaches locates it in a rank higher than 80. This trend is seen in most of the evaluations involving low quality single source candidate lists, without compromising the performance of the scenarios including high quality lists.

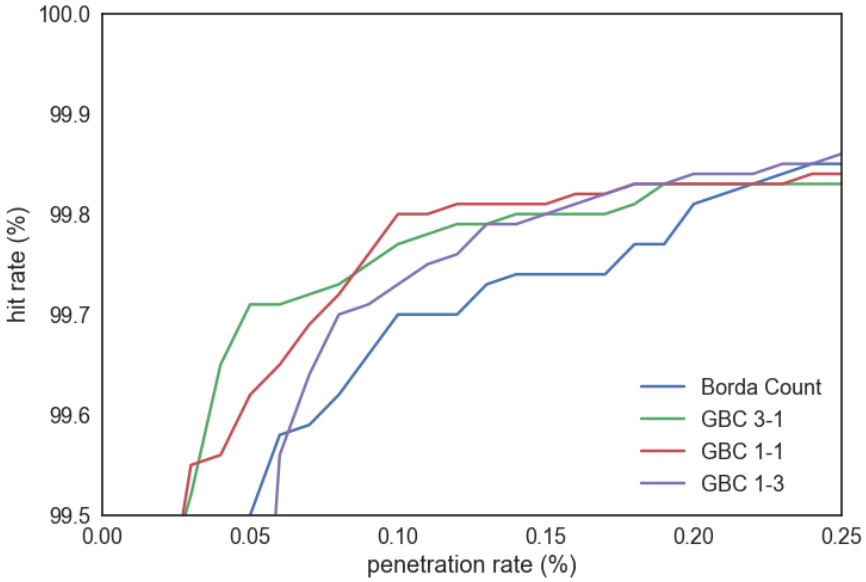


Figure 6.3: the effect of the cost ratio on indexing performance.

6.6 Summary

This chapter aimed at enhancing the efficiency of multi-biometric searches in large databases by proposing a generalized multi-biometric retrieval approach that aims at enabling an optimized inclusion of different biometric sources.

Driven by the need to replace the impractical exhaustive search in large databases, biometric data indexing presents a chance to limit this search to a small subset of the database. The utilization of multiple biometric sources in the indexing structure enables even higher retrieval performance, and thus a faster biometric search. This chapter proposed a rank-level approach that can be adapted to different modalities and single-source indexing structures. The flexible optimization possibility allows the integration of indexing approaches with varying behaviors. The proposed approach is also extended to include further information in the form of efficiently calculated approximate distances, rather than ranks. With the aim of presenting a solution that fits in realistic application, the proposed indexing solution was designed to maintain high level of accuracy when facing missing data or single-source candidate lists of low quality.

The studied case of multi-instance iris indexing (left and right iris) was chosen as both irises are usually captured simultaneously. Besides that, iris is one of the most accurate biometric characteristics and is already in use in large-scale biometric deployments. Iris codes were first transferred into a rotation invariant space using a novel transformation approach, then used to build an easily maintainable indexing structure based on LSH-Forest. This single iris indexing approach was used as the bases of the carried experiments where accuracy was measured for different levels of missing data and qualities of candidate lists.

Quality grade combination		$\phi = 0$		$\phi = 0.5$	
		$\mu = 0$	$\mu = 100$	$\mu = 0$	$\mu = 100$
++	++	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
++	+	0.18 ± 0.78	0.04 ± 0.18	0.17 ± 0.79	0.03 ± 0.18
++	0	2.53 ± 4.26	0.16 ± 0.37	2.5 ± 4.18	0.17 ± 0.37
++	−	8.03 ± 10.66	0.20 ± 0.40	8.66 ± 11.96	0.21 ± 0.41
++	− −	12.28 ± 14.20	0.16 ± 0.37	19.80 ± 20.69	0.27 ± 0.46
+	+	1.27 ± 2.54	5.34 ± 6.20	1.31 ± 2.57	3.22 ± 3.43
+	0	7.33 ± 7.38	11.22 ± 11.50	7.30 ± 7.40	6.64 ± 7.18
+	−	19.29 ± 15.35	11.36 ± 11.58	19.82 ± 17.06	9.79 ± 15.10
+	− −	31.21 ± 17.21	11.53 ± 11.71	42.50 ± 23.44	45.73 ± 25.52
0	0	19.90 ± 13.35	54.26 ± 10.68	19.88 ± 13.28	29.05 ± 13.98
0	−	43.49 ± 17.57	58.56 ± 10.35	35.63 ± 20.66	35.38 ± 19.87
0	− −	69.29 ± 6.80	60.25 ± 10.75	77.25 ± 6.77	83.54 ± 12.65
−	−	80.52 ± 16.45	88.94 ± 11.94	56.14 ± 24.30	53.21 ± 25.64
−	− −	96.32 ± 6.50	94.65 ± 9.11	97.48 ± 4.62	98.21 ± 4.78
− −	− −	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00

Table 6.4: multi-biometric indexing performance given by the rank (average ± standard deviation) of the correct identity in the retrieved list for different candidate lists (single source) quality combinations - GBC, different ϕ and μ.

Evaluation conducted on a database of 10k references and 10k probes of left and right iris images proved the validity of the proposed approach. This was demonstrated by surpassing state-of-the-art indexing accuracy and proving flexibility to missing data and low quality candidate lists.

This chapter responded to *RQ7* by proposing a generalized multi-biometric retrieval approach that enables an optimized and generalized inclusion of different biometric sources while maintaining high accuracy when facing missing data and low quality candidate lists. Next chapter will present a number of miscellaneous multi-biometric processes that complement the multi-biometric system work-flow.

Quality grade combination		$\phi = 0$		$\phi = 0.5$	
		$\mu = 0$	$\mu = 500$	$\mu = 0$	$\mu = 500$
++	++	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
++	+	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
++	0	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
++	−	0.88 ± 4.59	0.02 ± 0.13	1.35 ± 7.32	0.02 ± 0.13
++	− −	13.77 ± 13.07	0.22 ± 0.42	19.57 ± 19.36	0.20 ± 0.40
+	+	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
+	0	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
+	−	1.51 ± 6.56	0.03 ± 0.17	2.42 ± 10.76	0.03 ± 0.18
+	− −	22.51 ± 14.20	0.35 ± 0.48	33.63 ± 22.30	0.34 ± 0.47
0	0	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
0	−	1.64 ± 6.60	0.04 ± 0.20	2.86 ± 11.61	0.04 ± 0.20
0	− −	23.49 ± 11.31	0.46 ± 0.50	37.49 ± 19.48	0.46 ± 0.50
−	−	3.81 ± 11.60	0.45 ± 6.15	5.99 ± 17.04	0.43 ± 5.99
−	− −	30.36 ± 21.73	6.73 ± 24.19	43.27 ± 24.31	6.75 ± 24.19
− −	− −	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00

Table 6.5: multi-biometric indexing performance given by the rank (average ± standard deviation) of the correct identity in the retrieved list for different candidate lists (single source) quality combinations - GDBC, different ϕ and μ.

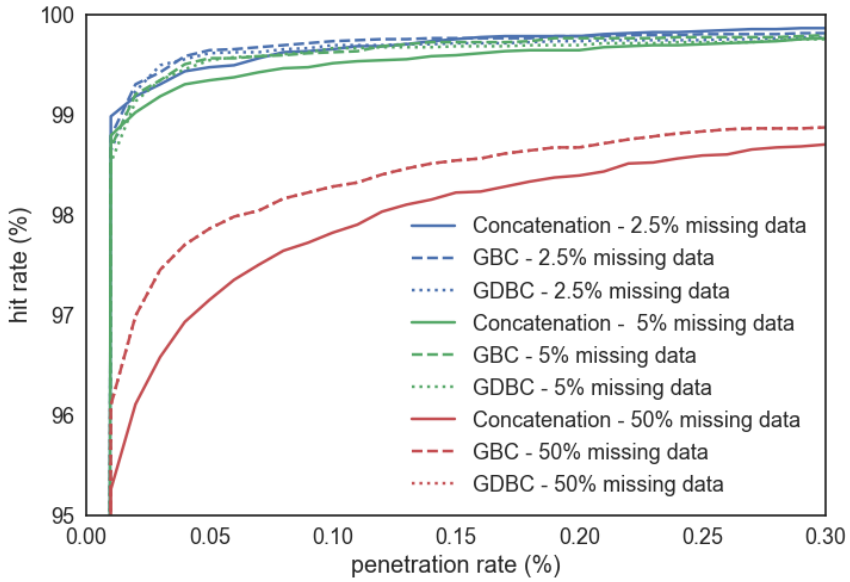


Figure 6.4: the effect of different levels of missing data on the indexing performance (GBC & GDBC cost ratio 3-1).

7 Miscellaneous multi-biometric processes

Previous chapters (3, 4, 5, and 6) discussed core operations in the multi-biometric work-flow. This chapter goes further by introducing different solutions that complement and utilize multi-biometric fusion. First, this chapter discusses the joint use of behavioral and physical biometric characteristics to assure a continuous author authentication. The second application deals with the practical use of face presentation attack detection. The third part discusses utilizing multi-biometric feature-level fusion to create more informative face reference templates from videos. This chapter is based on the publications [DMB16, DD16, DSN14]

7.1 Introduction

This chapter is concerned with three solutions that utilizes multi-biometric fusion to achieve better performance, efficiency, or enable new application scenarios. This section provides a brief introduction into the presented solutions.

7.1.1 Multi-biometric continuous authentication

Biometric recognition is typically used in conjunction with an access control (e.g. log-in) process. This means that the individual is recognized once at the start of a process, in order to get access to a system/service. In some scenarios, an attacker could gain access to the system after this initial log-in. One such scenario could be a stolen corporate laptop that the genuine user is still logged into.

Continuous authentication monitors the current user for the duration of the work session. Therefore, it can be used to protect from the aforementioned attacks. However, using continuous authentication also introduces some constraints to a typical biometric system. A genuine user with legitimate access should ideally not be interrupted during the working session. Therefore, biometric characteristics which require interactions with sensors, or otherwise interrupt the user during his work session, are not suited for continuous authentication systems. As a result, research has been focused on behavioral biometric characteristics such as keystroke dynamics, mouse movements and combinations of both of these biometric characteristics [BB09, MB13, DNN13].

In order to implement a continuous authentication, Bours introduced the concept of the trust model to continuous authentication [Bou12]. The trust model describes the confidence of the current user being the genuine user in the trust value. It also defines how the behavior of the current user affects this trust value.

This work aims at designing a trust model that can be adjusted and used to combine multi-biometric sources, a biological characteristic, and a behavioral biometric characteristic. Face recognition was chosen as the biological characteristic as it does not require additional interaction with a sensor, nor does it interrupt the work session of a genuine user. In order to minimize the impact on the privacy of the user, periodical pictures were captured from a webcam instead of a permanent video. For the behavioral characteristic, keystroke dynamics was chosen. Keystroke dynamics in continuous authentication is a well researched field [BW12, ZD15, BM15] and should complement the face recognition adequately.

Since the user is not always typing throughout the entire work session and the face recognition is performed periodically, changes to the trust value need to be made asynchronously. This requires an asynchronous information fusion approach.

A goal of this chapter is to develop and evaluate the feasibility and performance of a multi-biometric asynchronous continuous authentication system. It also aims at stating the main challenges facing the development and deployment of such a system in realistic conditions. As the bases for the face recognition sub-system, local binary linear discriminant analysis (LBLDA) [FR11] solution was used. For keystroke dynamics, a statistical method introduced by Bours et al. [BB09] was implemented. The work also introduces a specifically collected multi-biometric continuous authentication database.

7.1.2 Face presentation attack detection

Personal identification takes place in many domains, primarily as a mean of providing access control for security sensitive environments [SB14]. Conventional authentication methods depend on passwords or identity documents. However, these approaches proved to be easily spoofed and therefore do not meet the security demands of modern applications. Automatic biometric recognition algorithms utilize physical or behavioral characteristics to verify or identify individuals more securely [JRN11].

As promising and effective biometric solutions proved to be, they have become nonetheless subject to fraudulent attacks. Therefore, the vulnerability of such systems against fake biometric characteristics is a growing concern. The so called presentation attacks can be expressed in terms of, but not limited to, someone posing as another individual or hiding their identity [TLLJ10]. Subsequently, that led to the development of "presentation attack detection" PAD (or "anti-spoofing") techniques. Prerequisite to a good and reliable PAD application is most notably the ability to perform well with different kinds of attacks and scenarios under diverse conditions.

One of the most commonly accepted biometric characteristics is the face image. The main reason, among others, is the non-intrusive capture using non-contact sensors to capture it from a distance. Although face recognition is a task that the human brain can perform routinely and in an effortless manner, automated face recognition has been a challenging milestone in biometrics. In order to accurately identify a face, a myriad of factors have to be considered. Besides being invariant to age, pose, and facial expression, face recognition systems should take factors, such as varying illumination or changes due to accessories, into account [LJ11].

Unfortunately, as convenient as face biometrics has been established to be, recent works have shown that it is quite vulnerable to presentation attacks [MNL14]. In its most basic form, a face recognition system is designed to only recognize identities without concerning whether the subject is real or fake. Therefore, it can be easily spoofed by exposing the system's sensors to a printed photograph of the impersonated character [CAM12]. As a result to the growth of interest in face recognition systems, a number of different methods have since been developed to perform PAD.

This work aims at investigating the practical use of face PAD. This is achieved through three main aspects. The first is presenting an optical flow based PAD solution that proved to outperform the state-of-the-art works in most experiments. The second is to perform cross-database evaluation to simulate a more realistic scenario. This evaluation included listing a comparison with the latest published works. The third aspect is to analyze the duration (video length) required to achieve a confident decision, which is aimed at providing valuable information about the usability of such a system. The presented PAD solution was based on optical flow in a similar manner to *HOOF*-based feature extractor along with an AdaBoost [SS99] classifier. The cross database evaluation was carried on the *REPLAY-ATTACK* [CAM12], *MSU-MFSD* [WHJ15] and *CASIA-FASD* databases [ZYL*12].

7.1.3 Face reference from video

Face recognition is a very popular biometric modality that is used in a wide range of applications in areas like access control and unattended border control with satisfying recognition performances. Moreover, photographs, video material from cell phones, internet video and video material acquired by surveillance cameras are available as evidence material in many criminal investigations. This material can be analyzed for containing the faces of certain subjects of interest.

Especially in re-identification and law enforcement applications, the analysis of surveillance videos for face recognition is not a trivial task. Face images in these videos are not ideal for face recognition algorithms because of varying pauses and expressions. This goes both ways, for reference videos and for probe videos. If a face in two video sequences is to be compared, a computationally intensive cross comparison between all frames can be made, which is critical when analyzing a large amount of videos. To avoid this, creating a face reference template of the reference videos can reduce the computational effort and, if performed correctly, enhance the accuracy. This requires that an appropriate face image has to be automatically selected, and if more key-face images are selected, a singular reference template has to be created.

The third part of this chapter focuses on the creation of face reference model from multiple faces detected in a video sequence (multi-captures). This solution aims at improving the performance of face recognition based on one captured face image, as well as avoiding the high computational complexity of $N \times M$ comparison used to compare all faces across two sequences in video face recognition. Two challenges are dealt with in this work, the informative key-face selection and the effective face feature fusion. A video provides a big number of faces to use, a good selection for a limited number of key-face images paves the way for feature fusion to produce a discriminant face reference. Key frame selection was studied in the literature for general videos as well as in emotion recognition applications [DAGG11, GWL*13]. This work presents key-faces selection approaches based on inter-user variation, entropy, and detection confidence. The final face reference is created by binary feature fusion with different approaches investigated. The development and evaluation of the presented solution use the YouTube Faces database [WHM11].

The next sections in this chapter present a background overview, methodology description, and the related experiments and evaluation results for the three aspects of focus, multi-biometric continuous authentication (7.2), face presentation attack detection (7.3), and face reference from videos (7.4). A final discussion of the chapter contributions is presented in Section 7.5.

7.2 Multi-biometric continuous authentication

Biometric technologies are used to grant specific users access to services, data, or physical spaces. The access control is usually performed at the start of a session that spans over a period of time. Continuous authentication aims at insuring the identity of the user over this period of time, and not only at its start. Multi-biometrics aims at increasing the accuracy, robustness and usability of biometric systems. This section presents a multi-biometric continuous authentication solution that includes information from the face images and the keystroke dynamics of the user. A database representing a realistic scenario was collected to develop and evaluate the presented solution. A multi-biometric trust model was designed to cope with the asynchronous nature induced by the different biometric characteristics. A set of performance metrics are discussed and a comparison is presented between the performances of the single characteristic solutions and the fused solution. This is concluded by stating the major challenges facing the deployment of such a system in realistic conditions.

7.2.1 Related work

In many scenarios, like access control to a computer, user authentication is performed only during the initial log-in. This leaves room for attackers to gain access to the system after the initial authentication. Klosterman et al. identified six differences between person-authentication schemes in general and biometric-based authentication [KG00]. One of these differences is the fact that many biometric characteristics can be tracked continuously. Thus, the addition of continuous authentication can protect the system from attacks conducted after the initial log-in, and therefore greatly improve the security of the system. Solami et al. describe continuous authentication systems with five basic components [SBCI10]. These components are the subjects, sensors, detectors, biometric database, and decision module.

Yap et al. proposed a continuous authentication system for a computer running Windows XP [YSKR08]. In their scenario, the comparison score was computed in frequent time intervals. They offered two options if the comparison score would fall below a selected threshold. One was to freeze the system processes and the other was to freeze the input. After a successful re-validation of the user's identity, the freeze was lifted and work could continue. Klosterman et al. built a system where continuous authentication was used to enhance the security of computers running a Linux OS [KG00]. First, the users logged in normally via a virtual console. The continuous authentication system would then periodically take pictures using a webcam to verify that the user is still present and verify his or her identity. The results of this authentication attempts are added to an authentication log. After each authentication attempt, this log is scanned for authentication failures. The user is logged off if the threshold for such failures is exceeded.

Azzini et al. proposed a different idea by including a trust model [AMSS08]. Their system calculates a trust value, which is the basis for all decisions made by the system. The initial value is the comparison score of the initial authentication using fingerprint and face image. Then, face images are extracted periodically from a video camera. Depending on the comparison score, the trust value is either maintained or decreased. If the trust value falls below a certain threshold, the user is asked to input his or her fingerprint again. Niinuma et al. [NPJ10] proposed a continuous authentication framework that combines continuous authentication with conventional authentication. Furthermore, the framework updates the biometric reference templates every time the user logs in through the conventional authentication process.

Face recognition is a very popular biometric modality that is used with satisfying performances in a wide range of applications. This is due to its high universality, measurability, and acceptability. Some works dealing with uncontrolled face recognition used hand crafted image features such as scale-invariant feature transform (SIFT) [Low04] and local binary patterns (LBP) [OPH96]. Higher performances were obtained by combining more than one of those methods [WHT10]. The face recognition technology evolved from feature based approaches into appearance based holistic methodologies. Some of the well-studied techniques are the principle component analysis (PCA) [BHK97] and the linear discriminant analysis (LDA) [LPV03a].

In an effort to build face verification algorithms that are more robust to variations in facial appearances than traditional holistic approaches, researchers proposed the use of local appearance based face recognition. An example of such a method is the block based discrete cosine transform (DCT) that was shown to outperform similar holistic appearance based approaches [ES05]. Following the advances in local appearance based face recognition, Fratric and Ribaric proposed the use of local binary linear discriminant analysis (LBLDA) [FR11], which will be the base for the face recognition subsystem in this work.

Keystroke dynamics belong to the category of behavioral biometric characteristics and are used for the recognition of individuals based on their typing rhythm. Banerjee et al. [BW12] distinguished between systems using keystroke dynamics for biometric authentication and identification purposes in multiple categories. One of the most important factors to consider is the type of text the keystroke dynamics system in question examines. The

first type is static, also referred to as structured, text resulting in a keystroke dynamics system that is text dependent. The second type is free, also referred to as dynamic text.

A wide variety of data can be collected from keystroke dynamics and various features can be extracted from this data [BW12,ZD15]. This includes, press-to-press latency, release-to-release latency, release-to-press latency, trigraph, n-graph, key hold time, keystroke latency, pressure/force, total duration, and speed. Additionally some secondary features can be derived by further processing this information such as the minimum and maximum speed of typing, the mean and standard deviations of the features and the entropy [BW12]. Banerjee et al. divide keystroke recognition algorithms into four different categories [BW12], statistical algorithms, neural networks, pattern recognition, and search heuristics and combination of algorithms.

The system presented in this work uses free text keystroke dynamics. The extracted features are the key hold time and the release-to-press latency between keys. The methods used for feature extraction and comparison use statistical algorithms and are based on the work by Bours et al. [BB09].

Multi-biometrics tries to use multiple biometric information sources to enhance performance and to overcome the limitations of the conventional uni-modal biometrics. Such limitations are noisy data, low distinctiveness, intra-user variation, non-universality of biometric characteristics, and vulnerability to spoof attacks.

Information fusion is used to produce a unified biometric decision based on multiple biometric sources. The fusion process can be performed on different levels such as the sample-level, feature-level, score-level, and decision level. Simple approaches such as the sum rule score-level fusion proved to achieve high performance compared to more sophisticated approaches [RJ03]. The fused biometric sources can belong to different characteristics, algorithms, instances, or presentations.

Score-level biometric fusion techniques can be categorized into two main groups, combination-based and classification-based fusion. Combination-based fusion consists of simple operations performed on the normalized scores of different biometric sources. These operations produce a combined score that is used to build a biometric decision. One of the most used combination rules is the weighted-sum rule, where each biometric source is assigned a relative weight that optimizes the source effect on the final fused decision. The weights are related to the performance metrics of the biometric sources, a comparative study of biometric source weighting is presented by Chia et al. [CSN10] and extended later by Damer et al. [DON14a, DON14b]. Classification-based fusion views the biometric scores of a certain comparison as a feature vector. A classifier is trained to classify these vectors optimally into genuine or imposter comparisons. Different types of classifiers were used to perform multi-biometric fusion, some of these are support vector machines (SVM) [SVN07, GV00, DO14], neural networks [Als10], and the likelihood ratio methods [NCDJ08].

For conventional biometric recognition, or in cases where the identity of the individual is identified or verified only once, all the scores are available at the same time. In cases where not all the information is present at the same time, or is performed multiple times over a time-span, an asynchronous fusion needs to be performed. This work will be based on an asynchronous combination-based score-level weighted-sum fusion approach.

The concept of the ***trust model*** was discussed by Bours [Bou12]. The trust model describes the certainty that an individual is a genuine user, often expressed as the trust value, over time/actions. This means, the behavior of the current user is compared to the template of the genuine user. Based on each single action performed by the current user, the trust value is adjusted. If the trust value is too low, then the user is logged out. Also part of the trust model is the penalty and reward function, which defines the change of the trust value. This amount of change can be fixed or variable.

Mondal et al. propose a variable trust model where the change of the trust value is dependent on the comparison score of the current action performed by the user as well as the threshold value between penalty and reward, the width of the sigmoid for the penalty and reward function and the upper limit for the reward and penalty respectively [MB15].

7.2.2 Database

In order to develop and evaluate the proposed system, a database of biometric data simulating the application scenario was collected. The collected database consisted of keystrokes and facial images. To do this, a separate application was developed. Additionally, a declaration of consent and background information was handed out to the participants. The participants were asked to collect data amounting to approximately three work days, a minimum of 18 hours over a minimum of three different days. Otherwise, there were no limitations for participants as the goal was to collect data that closely matches their normal working behavior.

Face images: the images were captured by a webcam in an interval of 30 seconds. After starting the data collection program, the user was asked to adjust the webcam so the face will be in a central position.

Keystrokes: the algorithm for keystroke dynamics used in this work is based on the one introduced by Bours et al. [BB09]. The data collection followed a similar approach. As shown in Table 7.1 each time a key was either pressed or released, the following parameters were recorded:

- The time since the session started, in milliseconds.
- The current time (derived from the system clock).
- The date.
- The pressed key.
- The type of event.

Note that the pressed keys were not recorded plainly. When the participant locked the PC, no pictures or keystrokes were captured until the PC was unlocked again.

Event	Pressed key	Time	Date
KeyDown	13	454399	2015-07-14.13:53:13
KeyUp	13	454462	2015-07-14.13:53:13
KeyDown	75	464118	2015-07-14.13:53:22
KeyUp	75	464212	2015-07-14.13:53:22
KeyDown	79	464336	2015-07-14.13:53:22
KeyUp	79	464414	2015-07-14.13:53:23
KeyUp	78	464586	2015-07-14.13:53:23

Table 7.1: example of the keystroke dynamics collected data of an individual

The data of each user was saved anonymously by assigning it a randomly generated identity number. The collected sets of data will be referred by a number resulting from the order in which the references where created. The resulting database contains the data of 14 participants.

7.2.3 Methodology

This section describes the behavior of the proposed system. This includes the usage of the two different biometric recognition processes and their effect on the decision-making of the system, as depicted in Figure 7.1.

In Figure 7.1, TV stands for the trust value and the system additionally uses a timer and two thresholds:

- Timer t1: this timer measures the constant time between camera captures.
- Threshold 1 (TH1): if the trust value falls below this threshold, the user will be logged off.

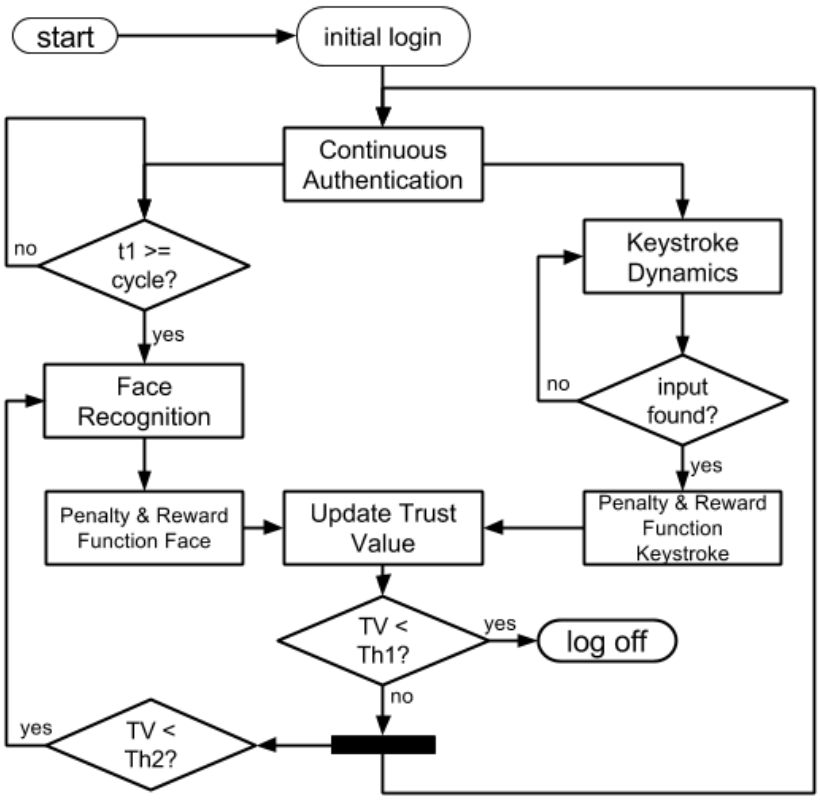


Figure 7.1: overview of the proposed multi-biometric continuous authentication system

- Threshold 2 (TH2): if the trust value falls below threshold 2, but is still higher than threshold 1, additional face recognition is performed outside of the timer cycle.

The key element of the system is the trust value. This value measures the confidence that the current user of the system is a genuine user. Certain actions will have an impact on the trust value, as they will increase or decrease it. If the trust value remains high, the user is believed to be genuine and can continue his/her work without interruption. If the trust value falls below a certain threshold (threshold TH1), the user is believed to be an imposter and is logged off. The trust value is determined as a function of the performances of the keystroke and face matchers, their resulting comparison scores, and the timer.

The keystroke dynamics is measured continuously. Each input is transformed and compared to the biometric reference saved in the database. The resulting comparison score is measured as the distance between the input and the reference and can have a big impact on the trust value, meaning that a high distance will decrease the trust value and a low distance will increase the trust value, depending on the deviation from the genuine behavior.

Face recognition is performed periodically. Using the face recognition only periodically lowers the negative impact on the users privacy and makes it harder to perform activity recognition compared to a permanent video surveillance. If the trust value falls below a threshold (threshold TH2), an additional attempt at face recognition is performed outside this periodic cycle. This gives an additional opportunity to increase the trust value of a

genuine user, preventing a false rejection in the cases where the trust value is decreased due to outliers. It also speeds up the rejection of imposters.

The effect of the face recognition on the trust value is handled in a similar way to the keystroke dynamics, thus the deviation from the threshold for a match has an impact on the loss of trust in the case of a mismatch. If no face is found in the capture, the trust value is decreased by a fixed value.

Scenarios: the following points outline typical scenarios to further explain the behavior of the system.

Genuine user scenarios:

- The user works normally on the PC. Keystroke dynamics are measured continuously and face recognition periodically as described above. Since there are deviations in the behavior of the user, the usage of both biometric characteristics should prevent or limit false rejections of genuine user despite these deviations.
- The user is logged in but leaves the PC. The periodic face recognition will not find a face in the image, thus the trust value will decrease. Since there are no keystrokes, the trust value will not increase. This will trigger the first threshold and therefore the additional attempt at face recognition. This attempt will fail too and the user is soon logged off, as the trust value falls below threshold TH1.
- The user works on the PC, but moves around a lot. This will possibly result in a failure to perform the periodic face recognition and thus a degradation of the trust value. The keystroke dynamics should keep the trust value above threshold TH1. If the periodic face recognition happens during a phase of inactivity and threshold TH2 is reached, a second attempt at face recognition is performed.

Imposter user scenarios:

- An imposter is working on the PC. Keystroke dynamics are measured continuously. Since the deviation of the behavior of the genuine user is bound to be quite high, the trust value will decrease fast. Face recognition will not be able to verify the user either, so the trust value will degrade even faster.
- An imposter is working on the PC, but moved out of sight of the sensor performing the face recognition. Since no face is found, the trust value will decrease. This will result in the same scenario as imposter scenario 1. This scenario requires the attacker to notice or to know about the face recognition sensor.
- An imposter is working on the PC, but moved out of sight of the sensor for the face recognition and is only using the mouse. Again face recognition will not be able to find a face and thus lower the trust value until the user is logged out. This scenario requires the attacker to notice or to know about the sensor for face recognition and the knowledge that keystrokes are used to verify the identity of the user.
- An imposter is working on the PC by only using the mouse (or as little keystrokes as possible) and has acquired a mask, or similar presentation attack tool to fool the face recognition. In order to deal with this, presentation attack detection (PAD) should be added to the face recognition subsystem. Otherwise, this might be a scenario in which the intruder might gain access to the system for a longer period of time depending on the quality of the face recognition versus the quality of the mask.

Trust model: the trust model consists of the reward and penalty function, which describes the behavior of the trust value and the range of the trust value. For all subsystems, the upper limit on the range of the trust value was set to 1. This limit aims at having a fast imposter rejection after an initial genuine user.

The lower limit of the trust value is equivalent to TH1. If the trust value falls below this point, the current user is logged out. This value was selected separately for each system and determined by testing (threshold at equal error rate).

Face recognition: a variable penalty and reward function was chosen for the face recognition. This means that the penalty and reward are not fixed but depend on the distance between the comparison score and the threshold, which decides if the current user is an imposter or the genuine user and therefore whether to punish or reward. Therefore, the effect on the trust value depends on the decision confidence (for both imposter and genuine decisions). The trust value is given by

$$TV = \begin{cases} 0 & \text{at startup} \\ \min(TV + (score - T), 0) & \text{if } score \geq T \\ TV - (T - score) & \text{if } score < T \\ TV - \gamma & \text{no face found} \end{cases}, \tag{7.1}$$

where *score* is the comparison score between the probe and the reference and T is the threshold for deciding whether to punish or reward based on the *score*. The threshold at the equal error operational point was used ($T = 0.59$). If no face was found in the capture, a penalty of $\gamma = 0.05$ was imposed on the trust value, this value has been determined by testing.

The threshold for rejecting a user has been set to -1 during the experiments. That means the range of the TV for face recognition is [-1,0].

Keystroke dynamics: for keystroke dynamics a similar penalty and reward function was used in a similar manner to the one proposed by Bours et al. [BB09]. This is a hybrid between a variable and a fixed trust model. The penalty to the trust value is calculated, while the reward to the trust value is a fixed value. Note that the reward has been increased in comparison to the implementation by Bours et al. [BB09]. This was done to cope with the realistic reference data where many keystroke combinations did not exist. The trust value is given by

$$TV = \begin{cases} 0 & \text{at startup} \\ \min(TV + R, 0) & \text{if } d < T \\ TV - (d + 0.3) & \text{if } d \geq T \end{cases}, \tag{7.2}$$

where d is the comparison score resulted from the keystroke dynamics expressed as the distance between a keystroke of the probe to the corresponding values of the reference, R is the reward and T is the threshold for deciding whether to punish or reward based on d. The following values have been found suitable, $T = 0.115$ and $R = 1.3$. The threshold for logging out a user has been set to -1 during the experiments. Therefore, the range of the TV for keystroke dynamics is [-1,0].

Fusion: since both comparison scores have an impact on the trust value but might not be always present at the same time or may be outdated, both penalty and reward functions from Equations 7.1 and 7.2 are part of the trust model of the fused system. As a result, every time the face recognition is triggered or a keystroke is inputted, the trust value is updated.

Score weights were introduced to control the effect of each subsystem on the final decision. These weights can increase or decrease the impact of the single components on the trust model. α was introduced as the weight for the face recognition and β was introduced as the weight for the keystroke dynamics. The *Timer* of the face recognition has been set to one minute. This means, the periodical face recognition occurs once every minute. The threshold for the additional face recognition steps has been set to $TH2 = \frac{TH1}{2}$. The overall trust value is given now by

$$TV = \begin{cases} 0 & \text{at startup} \\ \min(TV + \alpha * (score - T_1), 0) & \text{if } score \geq T_1 \\ TV - \alpha * (T_1 - score) & \text{if } score < T_1 \\ TV - \gamma & \text{no face found} \\ \min(TV + \beta * R, 0) & \text{if } d < T_2 \\ TV - \beta * (d + 0.3) & \text{if } d \geq T_2 \end{cases}, \tag{7.3}$$

where d is the comparison score output by the keystroke dynamic expressed as the distance between a keystroke of the probe to the corresponding values of the reference, R is the reward, T_1 is the threshold for deciding whether to punish or reward based on d, *score* is the comparison score of the comparison of the probe and the reference, T_2 is the threshold for deciding whether to punish or reward based on the *score*, γ is the amount of the penalty inflicted upon the TV if no face was found in the probe image. For these parameters the same values as in the uni-modal systems was kept, $R = 1.3$, $T_1 = 0.59$, $\gamma = 0.05$ and $T_2 = 0.115$. The lower range of the trust value in the fused system was set to -13 to cope with the realistic work condition were minimum typing and no frontal faces are available for a prolonged periods of time. α and β were set to 1 and 2 respectively.

7.2.4 Experimental setup

In order to conduct a sufficient evaluation of the proposed system and the used algorithms, the biometric data of each participant was divided into three parts. The whole biometric data of a participant is from now on referred to as a data-set. A subset of this data-set is a block.

The partition was done as follows: the amount of images belonging to a data-set was divided by three, this corresponds to the capture time. Each block contains one third of the face images. The boundaries of these blocks were then adjusted to coincide with recording sessions. That means, the boundaries were adjusted to match with the nearest start or end of a recording session respectively. The result are three blocks of face images, where the start and end of each block coincides with the start and end of a recording session. This was done to ensure that parts of one recording session would not be part of two different blocks. The timestamps of the start and end of each block of face images are then used to separate the keystroke data in blocks. As a result, the time-frame of a block of keystroke data is the same as the time-frame of the correlating block of face images. This resulted in three blocks of data for each data-set, where the blocks are separated by the start and end of a recording sessions.

For face recognition, an image from the first five images of each block was selected as a reference for each subject. The selection of the picture was based on pose and overall quality. One of the two remaining blocks was then used for the genuine test. Each block was selected as a reference once, and then compared to another block of the same data-set as a genuine test. Therefore, three genuine tests were performed per data-set. After that, each block of the other data-sets was compared to that reference as an imposter test. This made it possible to conduct a total of $14 * 3 = 42$ genuine tests and $42 * 39 = 1638$ imposter tests.

Performance metrics: in order to evaluate the performance of the algorithms discussed in this work, the following metrics where used:

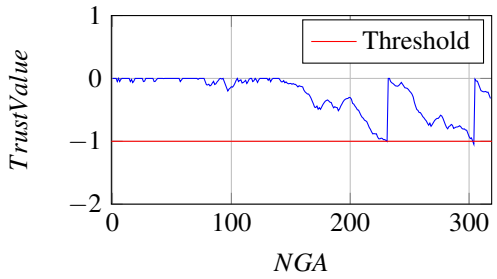
- *Imposter Detection Rate (IDR):* the rate of the successfully detected imposters.
- *Average Number of False Rejections (ANFR):* the ANFR indicates how many times the genuine user was falsely logged out during a session of a fixed period.
- *Average Number of Genuine Actions (ANGA):* the ANGA records how many actions a genuine user was able to perform on average before being falsely rejected by the system.

- *Average Number of Imposter Actions (ANIA):* the ANIA describes how many actions an imposter was able to perform before being detected and logged off by the system. Note that only the number of actions leading to the first detection and log off were recorded.

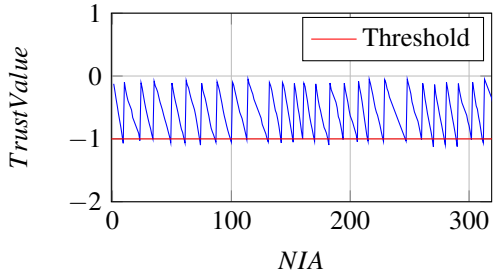
For keystroke dynamics, an action is a keystroke. For face recognition, an action is a face recognition comparison on an image taken by webcam. Consequently, it is easy to calculate the time the system needs to detect an imposter or the interval of false rejections from the ANIA/ANGA of the face recognition system. In order to compare these metrics more clearly, the average of the IDR, ANFR, ANIA, and ANGA over the users were calculated.

7.2.5 Results

Figure 7.2 shows an example of the development of the trust value over the number of actions for a genuine and an imposter user respectively based on face recognition. An action in this case is the periodical comparison of the probe to the reference. Thus, this can also be seen as a development over time as such comparison is performed every minute. Figure 7.3 presents an example of the development of the trust value over the number of actions for an imposter user and a genuine user based on keystroke dynamics. An action in this case is the input of a keystroke.



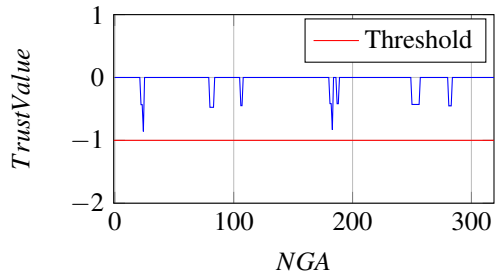
(a) Genuine user



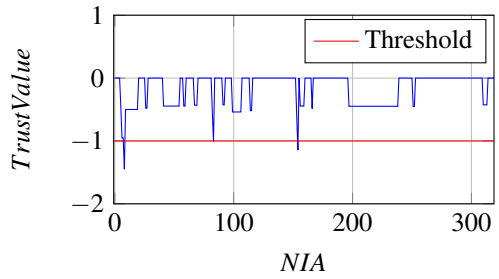
(b) Imposter user

Figure 7.2: an example of the trust value development based on face recognition over actions (time) for an imposter user and a genuine user.

An example of the trust value development over a working session based on the fusion approach is shown in Figure 7.4. Both the periodical comparison of the face recognition probe to the face recognition reference and the



(a) Genuine user



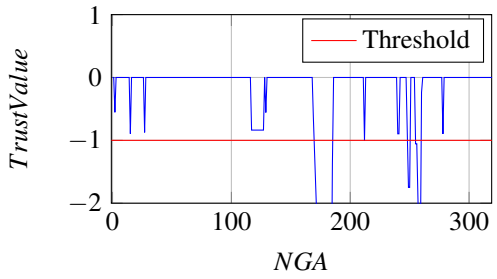
(b) Imposter user

Figure 7.3: an example of the trust value development base don keystroke dynamics over actions for an imposter user and a genuine user.

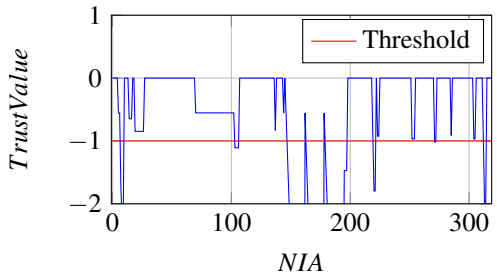
input of a keystroke are each considered to be an action, in order to monitor the development of the trust value. The three examples shown in Figures 7.2, 7.3, and 7.4 belong to the same user and the same working session.

The achieved performances of the fused solution is compared to these of face recognition and keystroke dynamics recognition in Figure 7.5. Two separate comparisons are presented because the definition of an "action" differs between the face recognition subsystem and the keystroke dynamics subsystem. When compared to the uni-modal face recognition system the fusion system performs better in the metrics ANFR and ANGA, while performing worse in the metrics IDR and ANIA, as seen in Figure 7.5. A similar pattern is noticed when comparing the fused solution to the uni-modal keystroke dynamics subsystem as seen in Figure 7.5. The results and the experiment procedure pointed out a number of challenges facing the design of such a system under realistic conditions. These challenges can be listed as follows:

- A large number of parameters are involved in designing such a system, tuning these parameters and measuring their effect on the performance is neither a direct, nor a trivial task. More effort should be made on creating direct links between these parameters and the operational requirements.
- A clear set of operational requirements have to be defined by the users or system integrators so it can be linked to the parametrization. This is clear for conventional verification systems, where simple error rates limits are specified (FAR and FRR). However, this is not the case in an asynchronous multi-biometric continuous authentication. This is still an open issue for future work.



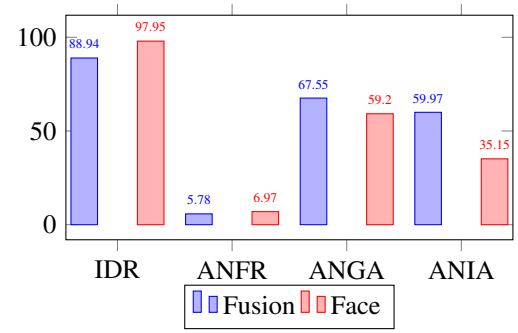
(a) Genuine user



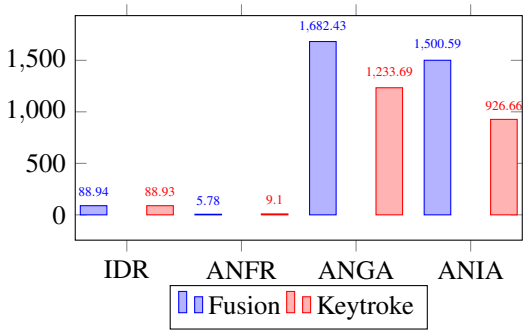
(b) Imposter user

Figure 7.4: an example of the trust value development based on the fused system over actions for an imposter user and a genuine user.

- Enrollment data for keystroke dynamics, as with other behavioral biometrics, is usually limited in comparison to the range of behavioral variations. For some users in the experiments of this work, there were very limited keystroke activities in very long periods, which affected the enrollment performance.
- Enrollment quality variation between users causes the system to behave differently with different users. To achieve a common set of security requirements, the quality of the enrollment has to be measurable and considered in the parametrization of the fusion system.
- By creating a realistic database with minimum restrictions, this work pointed out the challenging nature of the problem. A major step to tackle these challenges is to collect a larger, and more diverse, realistic data that enables deeper analysis of the problem.



(a) Fusion vs. face.



(b) Fusion vs. Keystroke dynamics.

Figure 7.5: achieved performance comparison between the fused solution and the single modality solutions.

7.3 Practical view on face presentation attack detection

Face recognition is one of the most socially accepted forms of biometric recognition. The recent availability of very accurate and efficient face recognition algorithms leaves the vulnerability to presentation attacks as the major challenge to face recognition solutions. Previous works have shown high preforming presentation attack detection (PAD) solutions under controlled evaluation scenarios. This section tries to analyze the practical use of PAD by investigating the more realistic scenario of cross-database evaluation and presenting a state-of-the-art performance comparison. The section also investigated the relation between the video duration and the PAD performance. This is done along with presenting an optical flow based approach that proves to outperform state-of-the-art solutions in most experiment settings.

7.3.1 Related work

Previous PAD works focused mainly on approaches based on texture and motion analysis of a 2D digital image. Texture based methods, such as the ones proposed in [CAM12] and [MHP11], take into consideration the different surface properties of human skin compared to a printed photo or electronic display. These differences reflect the information loss in those images, which in turns manifests itself in shape and detail loss. The employed techniques extract biometric features from single images captured with a dedicated camera. There are two essential feature extraction mechanisms to be distinguished, first, is Fourier transforming the 2D image into frequency domain and the second is applying the local binary pattern (LBP) operator. Frequency analysis, however, has been proven more error prone to images displayed on electronic screens in comparison to LBP. LBP based solutions show an overall half-total error rate (HTER) of 13.97% using the REPLAY-ATTACK database in combination with a support vector machine (SVM) classifier [CAM12].

Motion based approaches focus on the transition between consecutive images and thus rely on a frame sequence (video) in contrast to a single image. These techniques usually depend either on optical flow (as in [BLLJ09]) or feature comparison between frames (as in [AM11]) in order to detect specific movement patterns of a particular face part. These movements can be within the region of the eyes (as in [kJuJhY06]), the lips (as in [KFFB07]), or even the whole head. Some of these approaches require user collaboration, while others are completely non-intrusive. Focusing on optical flow based methods, Bao et al. ([BLLJ09]) exploited the fact that a planar object (a photograph) has a substantially different motion pattern compared to a 3D object (a real human face). That being said, it was demonstrated that such a system would be hard to spoof by a 2D image.

More recently, the approach proposed in [BDVS13] utilizes a histogram of oriented optical flow (HOOF) feature extraction technique, which in turns was initially introduced in [CRHV09]. Bharadwaj et al. report reaching a HTER of 1.25% using the REPLAY-ATTACK database in combination with principal component analysis (PCA) dimensionality reduction and linear discriminant analysis (LDA) as a means for classification [BDVS13]. However, as the authors of [WHJ15] recognized, the above mentioned application needs 230 frames in order to construct the descriptor (feature vector), which is consequently to be passed to the classifier. This circumstance renders the system not as user friendly and suitable for real-life scenarios.

7.3.2 Databases

As mentioned earlier, the presented solution is tested on multiple databases: the REPLAY-ATTACK [CAM12], the MSU-MFSD [WHJ15], and the CASIA-FASD [ZYL*12]. Each of these data-sets includes subsets for training and testing to evaluate the algorithm performance. These databases are used to evaluate intra and inter database performance.

The REPLAY-ATTACK database includes 1200 videos divided into three groups: 360 for training, 480 for testing, and another 360 used for threshold estimation. There were 50 identities involved in the recording process under two different lightning conditions: controlled (artificial lightning) and adverse (natural daylight). All the video clips were taken using the built-in camera of a 13″ Apple MacBook laptop with a resolution of 320 by 240 pixels at 25 frames per second and of 15 seconds (375 frames) each.

The spoofing attack videos can be in turn divided into two parts: in the first, the subjects display hard copies of high-resolution digital photographs printed on a plain A4 paper (photos taken using a 12.1 megapixel Canon PowerShot SX150 IS camera and printed using a Triumph-Adler DCC 2520 color laser printer). The second type of attack has the subjects displaying photos and videos taken with the aforementioned camera using an iPad screen with a resolution of 1024 by 768 as well as photos and videos taken with the 3.1 megapixel camera of an iPhone 3GS using its own screen. Each attack video is captured for about 10 seconds in two different attack modes: hand-based and fixed-support.

The MSU-MFSD database takes a similar approach as the REPLAY-ATTACK database but claims to generate better quality data using more advanced technology. It provides 280 recordings (only 280 are made publicly available out of 440) of 55 subject with 70 genuine and 210 spoofing attacks. The videos are taken with the built-in camera of a 13″ Apple MacBook Air laptop with a resolution of 640 by 480 pixels at 20 frames per second and of 10 seconds each as well as with the front-facing camera of a Google Nexus 5 device with a resolution of 720 by 480 pixels at 30 frames per second and of 15 seconds each. Thus, data is gathered with the aid of a mobile phone as well, simulating the application of mobile phone unlock.

The spoofing attempt videos can be divided again in two subsets. The first subset has subjects displaying hard copies of high-resolution digital photographs printed on plain A3 paper (photos taken using a Canon 550D SLR camera with a resolution of 5184 by 3456 pixels and printed using an HP Color Laserjet CP6015xh printer with a dots per inch of 1200 by 600). In the second subset, the subjects display high-definition videos taken with the aforementioned camera with a resolution of 1920 by 1088 pixels using an iPad Air screen as well as high-definition videos taken with the back-facing camera of an iPhone 5S with a resolution of 1920 by 1080 pixels using its own screen.

The CASIA-FASD database contains a set of 600 videos of 50 subjects, 150 videos of genuine faces and 450 of fake ones. The videos were captured with three different cameras. Two USB cameras with a resolution of 480 by 640 pixels. One of the USB cameras is brand new and the other is used for a long time to consider that long time usage degrades the image quality. The third camera is a higher definition Sony NEX-5.

What distinguishes the latter database from the former ones, is that the subjects are required to exhibit blinking behavior rather than keeping still during recording. As for the spoofing attempt videos, they could be divided into three categories. First, the subjects display the high-resolution genuine videos taken with the aforementioned Sony NEX-5 camera with a resolution of 1280 by 720 pixels using an iPad screen. The second has the subjects display hard copies of high-resolution digital photographs printed on copper sheets. While doing so, the attackers deliberately warp the photos, simulating a facial motion. In a similar scenario, the third attack has the subjects holding the photographs are now not warping them, but are rather required to exhibit blinking behavior through a cut off region on the photo. A variation of this spoofing technique is also utilized where the blinking process is simulated by moving an intact photo placed tightly behind a photo with a cut off eye region.

7.3.3 Methodology

The process of detecting presentation attacks is a typical binary classification problem. The motion-based approach used in this work can be summarized in the following steps. First, face detection is performed as most of the relevant information can be extracted from the area of the face and its close borders. Face detection followed

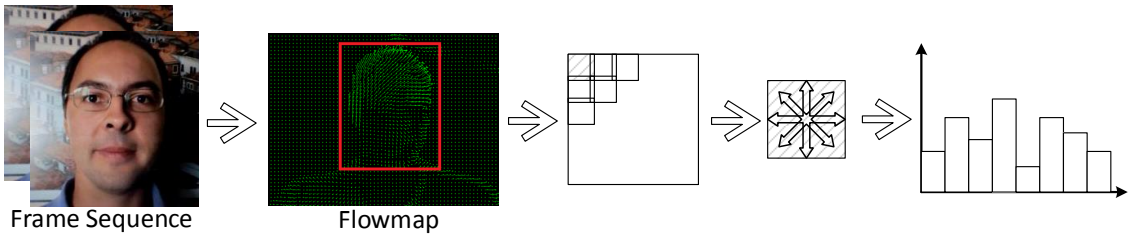


Figure 7.6: feature vector computation

the approach presented by Viola and Jones [VJ01]. Next, the optical flow map of the detected region of interest is calculated as described in [Far03]. Based on the dense optical flow map, a feature vector is calculated as explained in the next paragraphs.

Since the optical flow is based on the absolute pixel movement across consecutive frames, at least two frames are needed in order to begin further computation. However, the two frames needed to generate an optical flow map are not successive, since working at 20-30 frames per second does not guarantee a measurable pixel movement between consecutive frames. For this reason every second frame is skipped.

At each frame, face detection is conducted. The detected region of interest is rescaled to include a background area bordering the face. This area is important based on the assumption that the relative movement between the face and the background can be different between real and attack videos. To reduce the impact of different lightning conditions and standardize the feature vector calculation, histogram equalization and gray scale conversion were performed on the whole frame as well as rescaling the region of interest to 160x160 pixels. Detected faces in consecutive frames are neglected if their regions have an overlap value under 50%. This is done to avoid false face detections and in a more realistic scenario, a detection of a false face.

The optical flow is calculated between a pair of frames (skipping one frame in between) resulting in a flow map. The map is then divided into 24x24 pixel sized overlapping blocks grid (5 pixels overlap). To quantify the flows in each block, a histogram of all the enclosed optical flow vectors was created based on their directions. A histogram of 8 bins is used, where each bin including an optical flow direction range of 45° ($0° - 45°, 45° - 90°$, etc.). The values of the histogram are normalized (sum of histogram bins equal one). The histograms of all blocks are concatenated to form the feature vector (descriptor) representing the two frames. Figure 7.6 presents an overview of the feature vector calculation process.

In the following section, these feature vectors are used in three forms. First, as a single feature vector and is noted by F-single. Second, as a concatenated set of three consecutive (over time) vectors referred to here as F-triple. And finally, as a fused feature vector using feature values mean rule fusion for the set of three consecutive (over time) vectors, which is noted here by F-mean.

An AdaBoost classifier [SS99] is trained (for each experiment setting and each type of feature vector) to classify the vector into real or attack. AdaBoost classification was chosen to focus on more informative features in a pool of features that might contain many indistinctive elements. The classification returns a decision confidence measure based on the AdaBoost weak classifiers results, this confidence value is referred to here as the score. Score-level fusion was used to create a unified decision for longer portions of videos (videos that can create more than one feature vector). Simple combination fusion rules are used including the mean rule and the max rule [JNR05], noted here by S-max and S-mean respectively.

7.3.4 Experimental setup

The features vectors are created in three different approaches as described in Section 7.3.3. These feature vectors are calculated for the training data of each of the three used databases REPLAY-ATTACK, CASIA-FASD, and MSU-MFSD. For each database and feature vector type, an AdaBoost classifier is trained.

Cross-database evaluation was performed to simulate a more realistic PAD scenario. This is done as the PAD system in real use scenario is supposed to deal with varying video quality and different attack scenarios (unknown conditions). In the following, the experiments will be noted by the database used for training and the database used for testing, e.g. REPLAY-CASIA refers to an experiment setting where the REPLAY-ATTACK database was used for training and the CASIA-FASD database was used for testing. Evaluation within each single database was also conducted.

Previous works reported results of PAD assuming a single real/attack decision per video. In real applications, the time required to acquire such a decision is critical for the usability of a face biometric solution. Therefore, evaluation was also conducted on single descriptor performance. This results on evaluation of video segments of 3 frames (F-single) or 7 frames (F-triple and F-mean). An analysis of the performance improvement over the length of the video is also conducted and discussed in the next section 7.3.5.

Different evaluation metrics were used to be able to perform a comparison with the biggest possible number of previous works. Evaluation results are provided here as equal error rate (EER) and HTER values. Achieved true positive rate (TPR) values at fixed false positive rate (FPR) is also presented, true positive being a correctly classified presentation attack and a false positive is a bona fide presentation falsely classified as an attack. The FPR rates corresponds to the bona fide presentation classification error rate (BPCER) an the TPR corresponds to the complement of the attack presentation classification error rate (1-APCER). Both the BPCER and APCER were recently described in the ISO/IEC DIS 30107-3 [Int17]. Evaluation results are also shown as receiver operating characteristic (ROC) curves to view the tradeoff between the TPR and FPR for different thresholds. Here, the FPR and TPR are used as the conventional metrics in general binary classification problems, just as PAD. FPR and TPR can be viewed as a duplicate of *FAR* and *TAR* metrics more common in describing the verification performance of biometric systems (see Section 2.3.1). The evaluation of the proposed solution is consedered at the the PAD subsystem level as described in PAD subsystem as described in the standard ISO/IEC DIS 30107-3 [Int17].

The EER value is the common false negative rate (FNR) and FPR value at the threshold that makes both values equal. The HTER is the average value of the FPR and FNR at a certain decision threshold. The decision threshold utilized here is the threshold that produces the EER. The threshold value is obtained from the development database (not testing or training). In the experiment settings where the database did not contain a development set, the testing set was divided into subsets and the calculation of the threshold and HTER followed a cross validation approach. In these scenarios, the reported HTER value is the average value over these subsets.

7.3.5 Results

Figure 7.7a presents the ROC curves achieved by different experiment settings on the intra-database evaluation of the REPLAY-ATTACK database. It is clear that the decisions made over whole videos are more accurate than decisions made based on small part of the video (per descriptor, 3 or 7 frames). From these decisions based on single descriptor, concatenating three descriptors achieved better results than fusing them into a single vector using the mean rule. The performance based on whole videos seems to improve when using the score fusion mean rule, and as expected, when using fused information per descriptor (F-triple or F-mean).

The ROC curves achieved by the "F-triple" (selected as one of the best performing approaches) approach on different experiment settings are shown in Figure 7.7b, these results considered decisions made per descriptors (small portions of videos). The performance degradation is clear when performing cross-database evaluation. However, the cross-database performance improves when considering whole videos with score-level fusion as shown by the "F-triple: S-mean" approach in Figure 7.7c.

The required time to make a confident PAD decision is important for the usability of face recognition systems. Figure 7.7d presents the development of the PAD performance over the duration of the video. The results are shown for intra-database experiments on the "F-triple: S-mean" approach. It is clear that the performance improves rapidly during the first two seconds as a result of the score-fusion. One can notice that the performance reaches a saturation point after the first 3 seconds of a video sequence. The saturation duration is very similar for the three different databases.

Table 7.2 presents a wide comparison between the proposed approaches and the state-of-the-art works under different experiment scenarios. Different performance metrics were used to enable a wide range of comparison with published works. One can notice that the proposed approaches outperformed the state-of-the-art in most experiment settings, they were only outperformed significantly by the work of Wen et al. [WHJ15] when testing over the MSU-MFSD database (small improvement when testing on the CASIA-FASD database).

The authors of [BDVS13] showed that combining the HOOF feature with motion magnification achieved the best performance on the REPLAY-ATTACK database (HTER = 1.25%). However, motion magnification, cannot reach the reported performance without accumulating a large number of video frames (>200 frames), making these methods unsuitable for near real-time response [WHJ15].

From Table 7.2, one can notice the expected performance boost when considering whole videos. However, the performance does not always improve when using feature-level fusion (F-triple or F-mean). The score-level fusion mean rule usually outperforms the max rule as it has less focus on outlier scores.

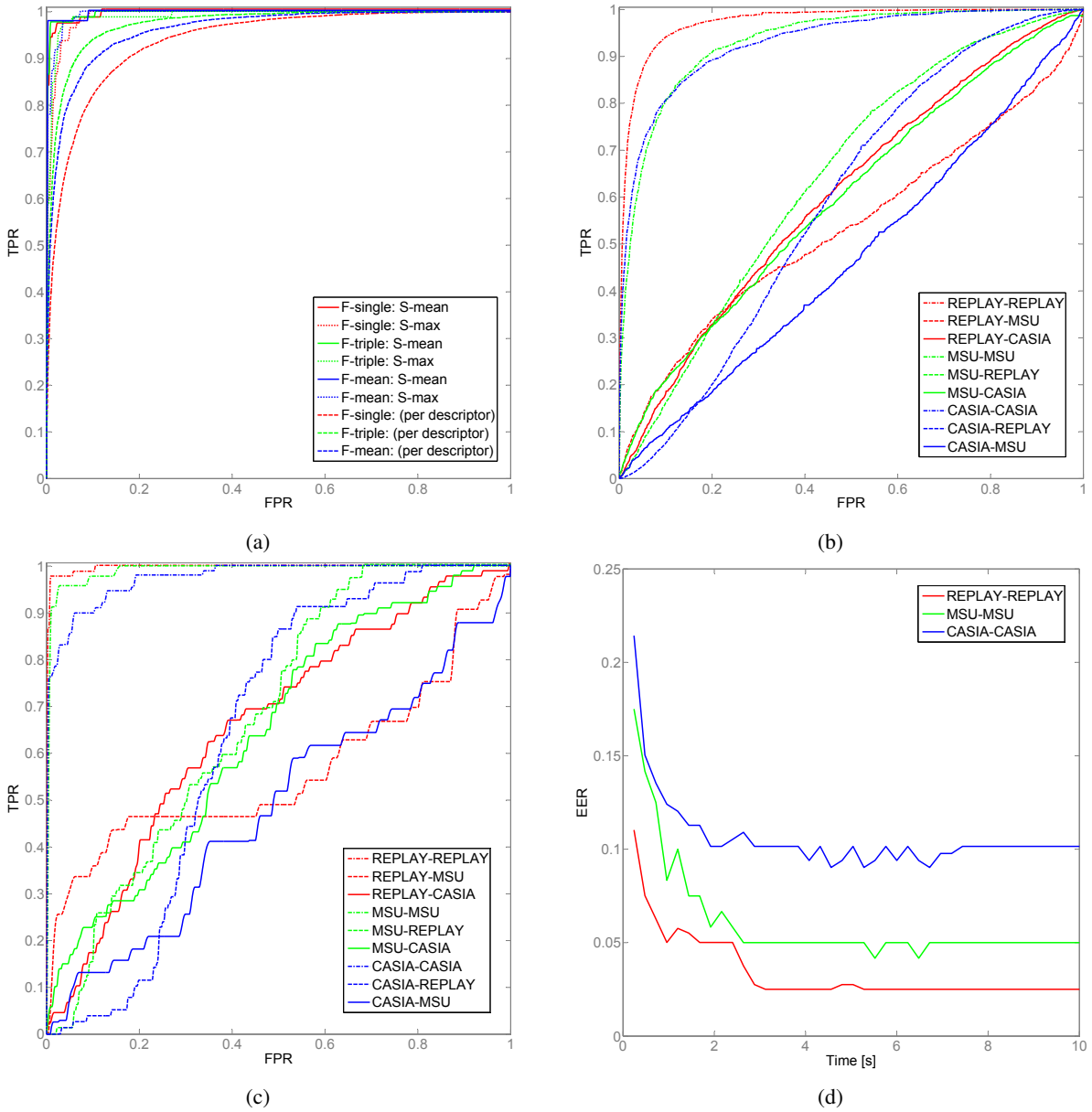


Figure 7.7: (7.7a) ROC curves achieved by different experiment settings trained and tested on REPLAY-ATTACK (7.7b) ROC curves achieved by the "F-triple" approach on intra- and cross-database evaluation. One classification decision per descriptor (3 or 7 frames). (7.7c) ROC curves achieved by the "F-triple: S-mean" approach on intra- and cross-database evaluation. One classification decision per video. (7.7d) The performance (represented by the EER) development at different video lengths (time). Evaluation on intra-database settings using the "F-triple: S-mean" approach.

	Train: REPLAY									**Train: MSU**									**Train: CASIA**								
	Test: REPLAY			**Test: MSU**			**Test: CASIA**			**Test: REPLAY**			**Test: MSU**			**Test: CASIA**			**Test: REPLAY**			**Test: MSU**			**Test: CASIA**		
	EER	HTER	TPR	EER	HTER	TPR	EER	HTER	TPR	EER	HTER	TPR	EER	HTER	TPR	EER	HTER	TPR	EER	HTER	TPR	EER	HTER	TPR	EER	HTER	TPR
[ZYL*12]: DoG+SVM	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	17	-	-
[CAM12]: LBP+χ^2	-	34.01	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[CAM12]: LBP+LDA	-	17.17	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	21.01	-
[CAM12]: LBP+SVM	-	15.16	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18.17	-
[CAM12] + [MHP11]: LBP + SVM	-	13.87	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	18.21	-
[BDVS13]: LBP+SVM	-	6.62	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[BDVS13]: HOOF+LDA+NN	-	**1.25**	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
[dFPAMM13]: LBP-TOP+SVM	8.17	8.51	94.5	-	-	-	-	61.33	-	-	-	-	-	-	-	-	-	-	-	50.64	-	-	-	-	21.59	23.75	82
[dFPAMM13, CAM12]: LBP+SVM	14.41	15.45	-	-	-	-	-	57.9	-	-	-	-	-	-	-	-	-	-	-	47.05	-	-	-	-	24.63	23.19	-
[dFPAMM13]: Corr [dFPAMM12]+MLP	11.66	11.79	-	-	-	-	-	48.28	-	-	-	-	-	-	-	-	-	-	-	50.25	-	-	-	-	26.65	30.33	-
[WHJ15]: IDA+SVM	-	7.41	92.2	-	-	**75.5**	-	-	-	-	-	73.7	8.58	-	92.8	-	-	**26.9**	-	-	-	-	-	**67.2**	12.9	-	86.7
[WHJ15, MHP11]: LBP+SVM	-	16.1	94.5	-	-	14.1	-	-	-	-	-	20.9	14.7	-	69.9	-	-	0.8	-	-	-	-	-	7.8	**6.7**	-	**93.3**
[WHJ15, KD12b]: DoG-LBP+SVM	-	11.1	92.1	-	-	19.5	-	-	-	-	-	23.6	23.1	-	62.8	-	-	4.6	-	-	-	-	-	14.2	12.7	-	84.4
F-single: S-mean (p.v.)	**2.51**	**2.75**	98.75	47.5	47.72	41	40.89	50.62	13.6	30.83	30.43	31.65	10	11.14	95	42.38	43.62	20.45	30.33	28.64	5.06	41.67	42.93	30.77	17.84	21.28	74.16
F-single: S-max (p.v.)	5.01	5.01	**100**	**30.83**	**30.51**	38.5	45.35	50.28	6.82	**25.31**	**24.44**	51.9	10	10.59	92.5	45.72	44.58	15.91	**28.32**	**28.43**	**31.65**	**35**	**35.62**	20.51	20.07	25.72	65.17
F-triple: S-mean (p.v.)	**2.51**	3	98.75	51.67	51.13	35.9	36.47	44.98	17.1	40.6	43.56	15.19	**5**	**5.52**	**97.5**	42.11	41.03	23.86	37.84	38.9	3.8	49.17	48.46	12.82	**10.15**	**9.92**	89.89
F-triple: S-max (p.v.)	3.51	3.63	98.75	45	45.07	25.6	36.47	41.44	10.2	45.36	39.14	**77.22**	8.33	7.24	95	**35.34**	**34.79**	23.86	35.34	35.57	10.13	54.17	54.11	12.82	**10.15**	10.12	**91.01**
F-mean: S-mean (p.v.)	**2.51**	4.26	**100**	50	50.87	25	**34.96**	**38.08**	23.6	47.62	46.89	10	**5**	8.43	95	42.86	41.9	24.72	37.84	38.99	2.5	45	44.84	10	11.28	12.52	88.76
F-mean: S-max (p.v.)	3.51	3	**100**	47.5	50.01	22.5	39.47	40.89	**25.8**	48.87	48.41	10	7.5	8.57	92.5	43.98	43.77	14.61	32.83	35.9	10	43.33	43.04	22.5	13.53	12.17	86.52
[CAM12] bob: LBP+SVM	17.43	16.98	-	45.67	44.67	-	-	-	-	-	-	-	27.89	26.84	-	40.89	40.73	-	48.14	40.73	-	36.1	35.97	-	23.77	23.27	-
F-single (p.d.)	13.48	13.47	82.17	44.81	44.76	21.5	45.64	46.99	12.7	38.24	38.56	22.33	19.66	19.85	64.9	44.91	44.82	16.18	40.81	41	14.76	45.24	45.23	19.8	26.41	27.88	52.68
F-triple (p.d.)	7.88	7.91	94.13	47.45	47.55	21.2	42.29	43.44	18	39.51	40.66	16.11	14.25	14.42	80.4	43.5	43.42	20.91	43.15	43.09	7.52	52.2	52.3	10.1	14.68	15.26	80.66
F-mean (p.d.)	10.31	10.28	89.55	50.69	50.86	14.1	40.97	41.42	19.2	47.28	47.01	10.34	20.16	20.26	65.7	43.3	43.41	16.91	43.3	43.59	8.16	46.88	46.6	13.68	18.64	19.04	70.88

Table 7.2: performance achieved on different experiment settings by the proposed approaches and state-of-the-art results (when available). TPR values are calculated at FPR value of 10%. All values are in percent(%), some are given per video (p.v.), others per descriptor (p.d.).

7.4 Face Reference from video: key-face selection and feature-level fusion

Face recognition from video in uncontrolled environments is an active research field that received a growing attention recently. This was mainly driven by the wide range of applications and the availability of large databases. This section presents an approach to create a robust and discriminant reference face model from video enrollment data. The work focuses on two issues, first is the key-faces selection from video sequences. The second is the feature-level fusion of the key-faces templates. The proposed fusion approaches focus on inducing subject specific feature weighting in the reference face model. Quality based sample weighting is also considered in the fusion process. The proposed approach is evaluated under different sittings on the YouTube Faces database and the performance gained by the proposed approach is shown in the form of EER values and ROC curves.

7.4.1 Related work

Some works dealing with uncontrolled face recognition used hand crafted image features such as SIFT [Low04] and LBP [OPH96]. Higher performances were obtained by combining more than one of these methods [WHT10]. The face recognition technology changed its focus from feature based approaches into appearance based holistic methodologies. Some of the well-studied techniques are the PCA [BHK97] and the LDA [LPV03a].

In an effort to build face verification algorithms that are more robust to variations in facial appearances than traditional holistic approaches, researchers proposed the use of local appearance based face recognition. An example of such a method is the block based discrete cosine transform (DCT) that was shown to outperform similar holistic appearance based approaches [ES05]. Following the advances in local appearance based face recognition, Fratric and Ribaric proposed the use of LBLDA [FR11] that will be discussed in more details in the following section.

As solutions based on deep learning recently dominated computer vision solutions, face recognition based on deep structures emerged with superior performances. Such a solution uses convolutional neural networks (CNN) feature extraction by combining a number of linear and non-linear operators. A representative example of such works is the DeepFace solution [TYRW14]. DeepFace utilizes a number of CNNs and a face alignment solution to bring the faces to a canonical pose using a 3D model. Further works by Sun et al. presented even better performing, CNN-based, face recognition solutions [SCWT14, SWT14].

The availability of large and suitable databases and evaluation protocols drove the advances in the field of uncontrolled face recognition. The main database used for such purposes is the labeled faces in the wild database (LFW) [HMBLM08]. More recently, a database with a similar structure was published with video sequences instead of images. This database, the YouTube Faces Database [WHM11], provides the opportunity to perform video recognition from video and multiple face fusion.

The need for a highly performing and robust on-the-fly face recognition for surveillance and access control applications drove the interest in face recognition from video. The availability of the YouTube Faces database allowed many researchers to develop innovative solutions for this problem. Wolf et al. presented the YouTube Faces database with a benchmark that compares a number of the available approaches and presented the match background similarity measure [WHM11]. Later on, Wolf and Levy presented an upgraded solution based on the SVM-minus classification [WL13].

In an effort to develop a pose variant face verification solution, Li et al. [LHL*13] proposed a probabilistic elastic method for face recognition. The proposed approach was applied and evaluated on face recognition from video task with satisfying results achieved. An approach using a local spatial-temporal descriptor based

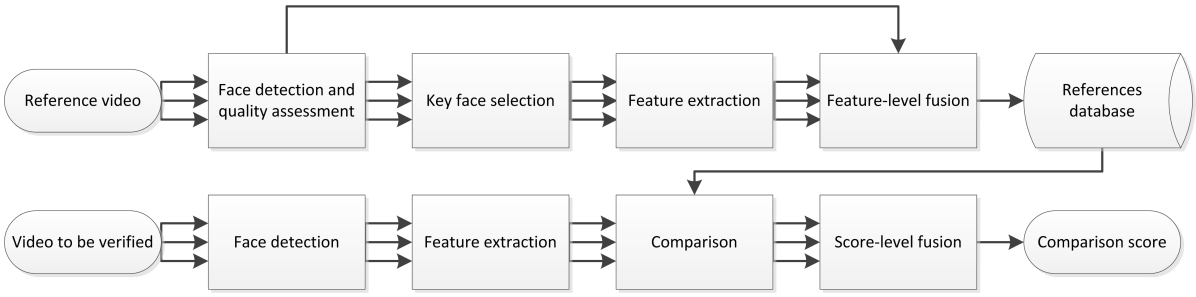


Figure 7.8: overview of the proposed face reference from video creation process and the overall face video comparison system.

on structured ordinal features was presented by Mendez-Vazquez et al. [VMC13] also aiming at improving the state-of-the-art in video face recognition. Dealing with the same problem, Cui et al. [CLX*13] tried to develop an alignment invariant solution based on regional spatial-temporal face descriptors.

Creating a face model from multiple face captures is performed by information fusion. Multi-biometric fusion combines biometric information from multiple sources taking in consideration certain weights that affect each source influence on the fused decision. The fusion process can be done on different levels such as, data [GBP04], feature [RDRK11], score [WTJ03], and decision levels [PJ01]. Multi-biometric sources can be classified into two broad groups, multi-modal biometrics and uni-modal multi-biometrics. The fusion process in multi-modal biometrics usually uses performance measures for source weighting such as the EER [CSN10]. Uni-modal multi-biometrics can be a result of multiple sensors, multiple algorithms, multiple samples, or multiple captures. In most cases of uni-modal multi-biometric systems, the fusion process is carried out on the data level [HST07] or on the feature level [PDC09].

7.4.2 Methodology

This section presents the proposed approach to create a robust reference for 2D-face images. This includes feature extraction and the proposed key-face selection and feature-level fusion approaches. The face video comparison process followed in this work is also discussed. Figure 7.8 presents an overview of the overall process.

Feature extraction: the features extracted from face images was based on the LBLDA initially proposed for face recognition solutions by Fratric and Ribaric [FR11]. To explain the functionality of LBLDA algorithm, one must start by explaining the LDA algorithm.

LDA is a machine learning method used to find a linear combination of features in order to separate one or more classes. This technique was often used as an appearance-based method in image-based biometrics, especially in face recognition [LPV03a]. The LDA algorithm deals with an image as a vector with the pixel values of the image as its elements. This algorithm transforms the images (vectors) into a space in which the between-class variance is maximized and the variance within classes is minimized.

The conventional LDA algorithm faces a computational problem when dealing with high dimensional feature vectors (i.e. number of pixels in an image) while having a relatively small number of classes as well as a low number of images per class. This was usually solved by using a dimensionality reduction technique such

as PCA [BHK97] or less conservative variations of the LDA algorithm such as the regularized discriminant analysis [LPV03b].

The LBLDA algorithm aims to combine the characteristics of appearance-based and local feature extraction methods. This is achieved by initially dividing the image into a set of overlapping sub regions. This division is performed using a sliding window over the image. This sliding window can be specified by size and step size to achieve different scales of sub regions and different degrees of overlap between these regions.

This work aims to deal with images taken under uncontrolled environment, as well as, being computationally efficient. Therefore, binary features (LBLDA) were chosen to be used. Binary features are believed to provide a higher level of measurements which offers robustness to image variations [FR11]. Moreover, binary features give the chance to perform faster and more efficient calculations.

Key-faces selection: to create a reliable face reference, a number of key-faces must be chosen from the enrollment video. These face images should be of a good quality and should represent the biometric face properties robustly and distinctively. In the following we propose three different criteria for key-face selection.

1. Face selection by entropy: entropy is a measure of the information content in data. Entropy is usually used as a measure of image quality [TLM08]. Based on that, one of the proposed approaches for key-faces selection is to choose the face images with higher entropy, i.e. higher quality and information content. The entropy of a face image I is calculated here by summing up the entropy of each of the three channels of the image. The entropy of each image channel is the sum of all pixel values probability ($p(i)$) multiplied by the second log_2 of these probabilities. The probability of a pixel value ($p(i)$) is obtained by calculating a normalized histogram of the possible pixel values (here, $i = \{1,\ldots,2^8\}$). The entropy of a 3-channel, 8-bit image can be formulated as

$$E(I) = -\sum_{C=1}^{C=3}\sum_{i=1}^{i=2^8} p(i)\log_2(p(i)). \tag{7.4}$$

The entropy values of all the faces detected in the video sequence of a subject are calculated. The faces with the highest entropy are chosen to represent the face of the subject. Figure 7.9a presents examples of face images with relatively high and low entropy.

2. Face selection by detection confidence: face detection in this work was based on a multi-scale sliding window approach. In this approach, a window moves over the search image with a certain step size and with different scales. At each position and scale the area covered by the window is classified into a face (positive) or a non-face (negative) area. This process followed the method presented by viola and Jones [VJ01], which used a cascaded series of AdaBoost classifiers to classify each sub-image into a face or non-face based on the extracted Haar-like features. The detection decision is insensitive to small changes in scale and transition of the sliding window, and thus, multiple detections usually occur around a detected face. These multiple detections are usually grouped to form the main face detection.

The number of detections grouped to from each final face detection is considered as the confidence of the detections [ZHHD16]. As the detector is trained on frontal faces, this confidence can be interpenetrated as a confidence of the face being frontal. The relative frontal face detection confidence between different video frames of one subject indicates the quality of the face image (less occlusion, frontal pose, and better illumination) and thus, faces with the highest detection confidence are considered as key-faces. Examples of faces with relative high detection confidence and low detection confidence are shown in Figure 7.9b. One can notice that the faces with high detection confidence tend to be more frontal and thus may contain more stable information.

3. Face selection of most different faces: one of the biggest influences on face recognition performance is the inter-variations in face appearance (of one subject) due to the changes in face expressions or the movements of

(a) (b) (c)

Figure 7.9: selected samples from the YouTube Faces database [WHM11] (a) Face samples with low (top) and high (bottom) entropy. (b) Face sample with low (top) and high (bottom) detection confidence. (c) Face samples of same subject with high difference (within one video).

the face. This face selection technique tries to find the most different faces of the subject of interest in a video sequence. This selection will be used later to minimize the effect of the face features that are not stable between these images, and thus focus more on the stable face features. The similarity measure considered to choose the most difference faces is based on the LBLDA approach discussed in Section 7.4.2. The key-face selection of different faces is performed as described in Algorithm 7.4.2.1.

Algorithm 7.4.2.1 select key-faces: most different faces

Given a set of faces F of one subject detected in different frames of a video, $F = \{f_1, \ldots, f_N\}$. Select a set of key-faces S that contains K faces as follows:

Initialize: select first key-face randomly and add it to the list S

for $i = 1$ **to** K **do**
select f_n from F so that it maximizes

$\max \sum_{\forall f_k \in S} Ditance_{LBLDA}(f_k, f_n)$

add the selected face f_n to the key-faces S
end for

An example of the most different faces of a subject from one video are shown in Figure 7.9c. These images show the peak difference between the face expressions of a certain subject, such as open and closed mouth or eyes. To benchmark the presented face selection approaches, key-faces were also selected randomly from the reference videos.

Feature-level fusion: feature-level fusion aims at creating a single feature reference vector out of the feature vectors extracted from the key-faces images. The fusion process is performed in order to create a more stable and robust representation of the subjects face. This is achieved by disregarding any irregularities that may occur in certain face images, and are not stable across the set of key-faces.

Given a set of key-face images X where $X = \{x_1, x_2, \ldots, x_N\}$ of one subject. Each image x_n can be represented as a binary vector B_n using the LBLDA approach. A binary vector B_n contains K binary values that represent the face image x_n, $B_n = \{b_n^1, b_n^2, \ldots, b_n^K\}$. A binary feature b_n^k can take one of two values $\{-1, 1\}$. The task of the feature-level fusion is to create informative and discriminant reference feature vector R that represents the face of the subject in the key-face images set X. The vector R is created by combining the set of binary feature vectors SB of the selected key-faces, $SB = \{B_1, B_2, \ldots, B_N\}$.

Three main fusion approaches were presented and evaluated in this work. First is the *binary voting* that produces a binary reference feature vector out of a set of key-faces. The second and the third discussed approaches considered the stability of each feature across the key-faces. This stability is represented by the value given to the feature. This value acts like a personalized (related to the subject) weighting of each of the elements of the fused reference feature vector. The second and the third fusing approaches are referred to in the following as the *voting fusion* and *discriminant voting* Fusion.

1. Binary voting: given a set of binary vectors that corresponds to a set of key-face images, binary voting combines these vectors and produces a fused reference binary vector. Any element r_k of the fused binary vector $R = \{r_1, r_2, \ldots, r_K\}$ can be given by

$$r_k = \begin{cases} 1 & if \sum_{n=1}^{N} b_n^k >= 0 \\ -1 & if \sum_{n=1}^{N} b_n^k < 0 \end{cases} . \tag{7.5}$$

Each of the binary elements in the resulted fused reference represent the most stable binary case of this element across the vectors extracted from the key-faces.

2. Voting: using this approach, the element sign in the fused reference vector is determined such as in binary voting. However, the values of the fused features are not binary and they depend on the stability of these features within the set of the subject key-faces.

$$r_k = \begin{cases} \frac{\sum_{\forall_n, b_n^k >= 0} b_n^k}{N} & if \sum_{n=1}^{N} b_n^k >= 0 \\ \frac{\sum_{\forall_n, b_n^k < 0} b_n^k}{N} & if \sum_{n=1}^{N} b_n^k < 0 \end{cases} . \tag{7.6}$$

This approach integrates personalized feature weighting by assigning higher value to the reference elements that are more stable across the binary feature vectors of the key-faces.

Inducing weighting on the sample level is assumed to improve the performance by regulating the relative effect of the different key-faces based on their quality. Quality measures and weighting approaches used are discussed later in Section 7.4.2. The voting approach with sample weighting produces the fused reference vector element r_k as follows

$$r_k = \begin{cases} \frac{\sum_{\forall_n, b_n^k >= 0} w_n * b_n^k}{N} & if \sum_{n=1}^{N} b_n^k >= 0 \\ \frac{\sum_{\forall_n, b_n^k < 0} w_n * b_n^k}{N} & if \sum_{n=1}^{N} b_n^k < 0 \end{cases} , \tag{7.7}$$

where w_n is the relative weight of the sample (key-face) x_n.

3. Discriminant voting: just as in *voting*, the signs of the resulted fused reference elements are decided by the corresponding elements of the fused binary vectors. However, the effective value of the fused elements, which corresponds to their weight, are calculated differently. The fused elements here are calculated as follows

$$r_k = \frac{\sum_{n=1}^{N} b_n^k}{N}. \tag{7.8}$$

This approach insures lower weight for unstable features when compared to the previously discussed *voting* approach. Here, the stable and unstable feature elements are more discriminant and the difference between their values is larger.

Sample weighting can be also induced on *discriminant voting*. By introducing sample weights w_n, equation 7.8 can be rewritten as

$$r_k = \frac{\sum_{n=1}^{N} w_n * b_n^k}{N}. \tag{7.9}$$

Weighting: relative weights are assigned to each key-face image to control its effect on the final fused reference vector. The weights are calculated based on the quality measures of the detected face images. In this work, the entropy of the image (as in Equation 7.4) and the face detection confidence are considered as the quality measures. The weight of each image w_n is then used when fusing the feature vectors of different images such as in Equations 7.7 and 7.9.

The relative weight of an image x_n within the set of key-face images X based on the entropy is calculated as

$$w_n^e = \frac{E(x_n)}{\sum_{i=1}^{N} E(x_i)}. \tag{7.10}$$

When taking the detection confidence as the base quality measure, the relative weight of the image x_n is formulated as

$$w_n^{dc} = \frac{DC(x_n)}{\sum_{i=1}^{N} DC(x_i)}, \tag{7.11}$$

where $DC(x_n)$ is the detection confidence of the face in image x_n and is represented by the number of neighbors forming the final merged detection.

Comparison: comparing a face image to a subject reference (result of the feature fusion) results in a similarity score that indicates the degree in which the face image belongs to the subject of interest. Given a binary vector B extracted from a face image by LBLDA and the fused reference vector R calculated as described in Section 7.4.2, the similarity score S can be calculated as

$$S(R,B) = \frac{\sum_{\forall_k, \mathrm{sgn}(b_k)=\mathrm{sgn}(r_k)} |r_k|}{K}, \tag{7.12}$$

where the vectors R and B are of the length K.

7.4.3 Experimental setup

The development and evaluation of the proposed solution were performed using the YouTube faces database [WHM11]. This database provides the opportunity of simulating the realistic use case scenario by having uncontrolled face images in different situations, as well as providing the ability to perform face in video recognition using multiple face fusion.

The database was split into ten folds to perform cross validation, each split contained 250 genuine and 250 imposter comparison pairs. In each step of cross validation, nine of the folds are used for training while the last fold is used for evaluation. The results achieved are presented as EER and ROC curves, both as average over the

ten folds. The ROC is produced by calculating the false acceptance rate (FAR) and the false rejection rate (FRR) at each possible threshold value that separates genuine and imposter users.

The experiments were split into four parts based on the key-faces selection approach used (selection by entropy, detection confidence, most different faces, and random selection). For each of the four cases, eight different approaches were used to create a face reference out of the selected key-faces (ten faces in the reported experiment results).

The first of the eight approaches to create a reference is implemented to benchmark the results. This approach, noted as *No Fusion*, uses only one key-face (top entropy, detection confidence, or random) and thus does not depend on feature-level fusion. The other approaches for the creation of face reference from video are based on the feature fusion techniques discussed in Section 7.4.2 and the sample weighting techniques presented in Section 7.4.2. These approaches are binary voting, voting, entropy weighted voting, detection weighted voting, discriminant voting, entropy weighted discriminant voting, and detection weighted discriminant voting.

The results are also evaluated with an NxM comparison where all faces from both paired videos are compared to each other and a simple score-level fusion is applied to produce a final score. The results of this approach are considered as a base benchmark for the other experiments to show the positive effect of the proposed feature-level fusion, besides the clear efficiency advantage.

Each of the created references (one for each reference video) is compared to the paired video. The comparison is done with every face (frame) in the paired video sequence. The comparison with all the frames of the video resulted in a set of similarity scores for each video. These scores are then fused by simple combination rules. The combination rules used in this experiment are maximum rule (*max*), minimum rule (*min*), median rule (*median*), and mean rule (*mean*), which corresponds to the same performance as the sum rule. The combination of the scores resulted in one fused score for each video-video comparison.

7.4.4 Results

Tables 7.3, 7.4, 7.5, and 7.6 presents the EER achieved by the four different key-faces selection approaches, the eight different reference creation approaches, and the different score fusion rules. The results are also shown for the cross video $N \times M$ comparison, where all the frames from both videos are compared to each other and then fused with simple fusion rules. It can be noticed that the results produced by references created by feature-level fusion of key-faces are generally better than the performance of references produced by only one face image (*No Fusion*).

In general, fusing by voting and discriminant voting scored lower EER values when compared to the binary voting, this can be explained by the personalized feature weighting induced by these fusion approaches. Detection weighted discriminant voting scored best EER values when applied on key-faces selected by entropy. However, when faces are selected by detection confidence or by most different faces selection, the voting approach slightly outperforms the other weighted voting and discriminant voting approaches. It must be mentioned that recent works scored lower EER values on the same database [VMC13, CLX*13, LHL*13]. However, this work does not aim at achieving the highest performance regarding video face recognition but focuses on proving the sanity of creating face references by feature-level fusion that outperform references from a single capture and provide a more efficient solution in the case of video face recognition.

Score-level fusion combination rules have obvious effect on the results when compared to considering only one face from the video sequence, *first face*. *Median* and *mean* combination rules performed relatively better than other combination rules. The best scored EER value was achieved by selecting key-faces by most different faces selection, combining them on the feature-level by voting, and combining the resulting scores by the median score combination rule. Selecting key-faces randomly scored comparable results when fused by median fusion

rule. However, the degradation in the result with respect to other key-face selection approaches was clear in other score fusion methods.

It must be noticed that using Feature-level fusion to create face references increases the efficiency of video face recognition. While performing a cross video comparison requires $N \times M$ comparisons and a score fusion process of the same number of scores, video face recognition based on the proposed feature-level fusion approach requires only M comparisons and a score-level fusion process of M scores, given that N and M are the number of frames in the reference and probe videos respectively.

	No Fusion	Binary Voting	Voting	Discrimin- ant Voting	Entropy Weighted Voting	Entropy Weighted Discriminant Voting	Detection Weighted Voting	Detection Weighted Discriminant Voting	NxM score fusion
max	0.3392	0.3246	0.3256	0.3159	0.32	0.3153	0.3192	0.3150	**0.309**
min	0.3764	0.3624	0.3565	0.3604	0.3608	0.3599	0.3608	**0.3519**	0.390
median	0.3355	0.3233	0.3171	0.3182	0.3221	0.3182	0.3185	**0.3153**	0.316
mean	0.3384	0.3233	0.3203	0.319	0.32	0.3203	0.3208	0.3187	**0.316**

Table 7.3: face selection by entropy: EER values achieved.

	No Fusion	Binary Voting	Voting	Discrimin- ant Voting	Entropy Weighted Voting	Entropy Weighted Discriminant Voting	Detection Weighted Voting	Detection Weighted Discriminant Voting	NxM score fusion
max	0.3355	0.3249	0.3164	0.3192	0.3225	0.3188	0.3208	0.32	**0.309**
min	0.372	0.3642	**0.3517**	0.3629	0.365	0.3619	0.3653	0.3619	0.390
median	0.3291	0.3192	**0.3112**	0.3143	0.3165	0.3148	0.3169	0.314	0.316
mean	0.3275	0.3225	**0.314**	0.3167	0.32	0.3166	0.3182	0.3159	0.316

Table 7.4: face selection by detection confidence: EER values achieved.

	No Fusion	Binary Voting	Voting	Discrimin- ant Voting	Entropy Weighted Voting	Entropy Weighted Discriminant Voting	Detection Weighted Voting	Detection Weighted Discriminant Voting	NxM score fusion
max	0.3434	0.3218	0.3176	0.3211	0.3236	0.3223	0.3223	0.3185	**0.309**
min	0.3769	0.3656	**0.359**	0.3616	0.3632	0.3619	0.3645	0.3691	0.390
median	0.3332	0.3174	**0.3086**	0.3148	0.3172	0.3148	0.3187	0.3174	0.316
mean	0.3363	0.3208	**0.3153**	0.3187	0.3198	0.3182	0.3208	0.3187	0.316

Table 7.5: face selection of most different faces: EER values achieved.

The plot in Figure 7.10 presents ROC curves produced by different experiment settings. The ROC curves show the clear advantage of feature-level fusion used to create a single reference vector from the reference video. It also indicates the performance gained by the key-face selection method proposed with respect to selecting key-faces randomly, under the same score-level fusion rule.

	No Fusion	Binary Voting	Voting	Discrimin- ant Voting	Entropy Weighted Voting	Entropy Weighted Discriminant Voting	Detection Weighted Voting	Detection Weighted Discriminant Voting	NxM score fusion
max	**0.309**	0.3616	0.3614	0.3611	0.3606	0.3616	0.3616	0.3611	**0.309**
min	0.3857	0.3725	0.3722	0.372	0.3737	0.3722	**0.3717**	0.3725	0.390
median	0.3153	**0.3145**	0.3163	0.3156	0.3150	0.3158	0.3151	0.3161	0.316
mean	0.3207	0.3662	0.3684	0.3679	0.3681	0.3684	0.3692	0.3687	**0.316**

Table 7.6: face selection randomly: EER values achieved.

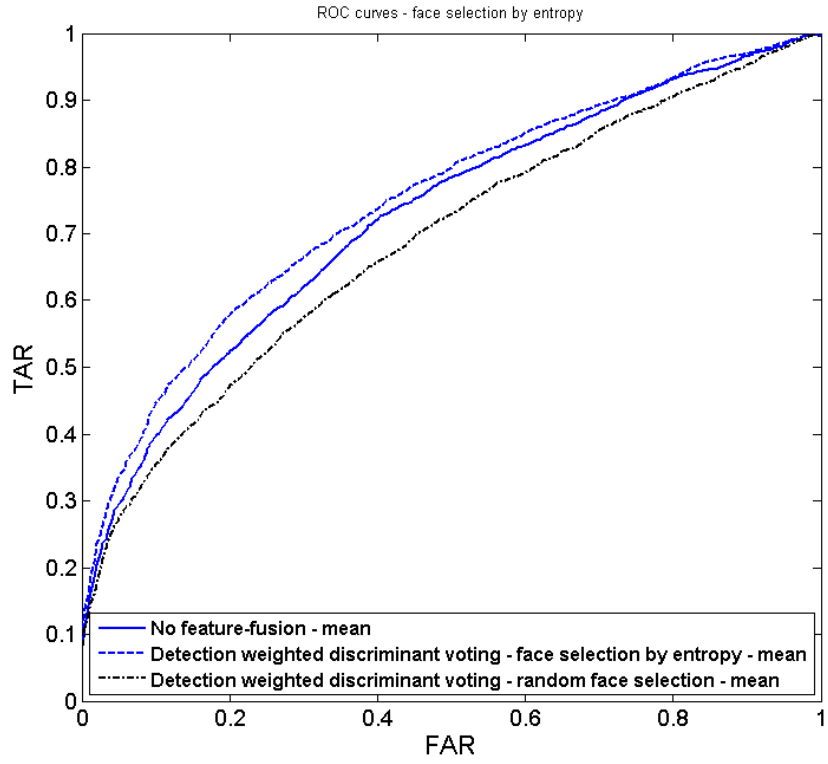


Figure 7.10: comparison between ROC curves obtained with/without feature fusion and with different key-face selection approaches.

7.5 Summary

This chapter presented three miscellaneous multi-biometric processes that utilize multi-biometric fusion to enhance its performance and enable new application scenarios. At first, this chapter discussed continuous authentication and presented a design of a multi-biometric, asynchronous, continuous authentication system. Later, face PAD under realistic application scenario was studied as an essential step of any biometric system, including multi-biometrics. Lastly, to enhance the accuracy and efficiency of face comparison between video sequences, key-face selection and feature level fusion were discussed.

Interest in continuous authentication is increasing as a tool to secure working session in security-critical applications. As in the more conventional biometric systems, multi-biometrics is expected to add flexibility and security to a continuous authentication system. This chapter presented a multi-biometrics continuous authentication that includes behavior (keystroke) and physical (face images) characteristics asynchronously. To develop and evaluate the solution, a realistic database was collected for the considered scenario. A multi-biometric trust model was designed to cope with the asynchronous nature induced by the different biometric characteristics. Results presented a direct comparison between the fused continuous authentication solution and the single characteristic solutions. This presented an answer to *RQ8* by proposing an initial design for such a system and proving that a similar solution can increase security and usability by reducing forced log-offs while maintaining high imposter detection rates in comparison to uni-biometric solutions. This was concluded by stating a set of challenges facing the development and deployment of such a system in realistic conditions.

In the second part of this chapter, vulnerability to presentation attacks was discussed as the main challenge preventing face recognition solution form being the main tool of identity authentication in a large number of applications. A number of previous works discussed presentation attack detection solutions without considering the capture period required to achieve confident decisions. Other works showed good performances in intra-database evaluations but failed to do so in the more realistic cross-database tests. Section 7.3 presented a comparison of the recent works considering cross-database presentation attack detection. An optical flow based approach that utilizes different fusion techniques was suggested and proved to outperform the state-of-the-art results in most experiment settings. The work also discussed the relation between achieving a confident presentation attack detection decision and the required capture time. As an answer to *RQ9*, this chapter pointed out weak performance points in the state-of-the-art algorithms when dealing with realistic scenarios. It also emphasized the role of information fusion in enhancing the accuracy of such a system and commented on the capture period needed to saturate the performance of such a fusion process.

Creating a face reference template from a video sequence captured in uncontrolled environment is essential to enable efficient face comparisons in surveillance scenarios. Creating such a template requires the automatic selection of diverse and representative captures of a subject face from a video sequence. It also requires a dependable feature-level fusion approach to create this single template from the selected key-faces. The third part of this chapter presented a novel approach for creating a 2D face reference model from video enrollment data. The reference model created by the feature-level fusion aimed at being robust and discriminant by considering the stability of features and inducing subject specific feature weights. The fusion process also considered sample weighting of the key-face images to control their relative effect on the final model. Different approaches were proposed to perform appropriate key-faces selection. The results were reported on the YouTube faces database with clear advantage of the proposed fusion and key-face selection approach on the face recognition performance, especially from the computational efficiency perspective. This proposed a response to *RQ10* by suggesting the use of face selection by selecting the most diverse faces along with a voting-based feature-level fusion, which proved to outperform an exhaustive comparison under simple mean rule fusion.

Previous chapters discussed the proposed solutions that aimed at enhancing different components of the multi-biometric work-flow. The next chapter will conclude the thesis and provide a brief outlook on future work.

8 Conclusions and future work

After the detailed responses to the research questions (Section 1.2) in the previous chapters (3, 4, 5, 6, and 7), this chapter presents a summarized conclusion of this thesis and an outlook for future research.

8.1 Conclusion

The use of multiple biometric sources within a unified frame, i.e. multi-biometrics, aims at circumventing the limitations of single source biometrics and thus enabling a wider implementation of biometric technology. This work presented advances in multi-biometrics by addressing different aspects of the multi-biometric system work-flow. These aspects were, pre-fusion processes, optimizing the fusion process by source weighting and integrating supplementary information, multi-biometric reference retrieval, and further processes that utilize and complement multi-biometric fusion.

Within pre-fusion processes, two topics were tackled. First is the possibility of building a link between the comparison score values and their induced performance through the normalization process. Here, as a response to *RQ1*, Section 3.3 proposed modifications to three widely used normalization techniques by anchoring a certain performance related point in the scores distributions. These performance anchored normalization approaches aligned score distributions from different sources at this performance related point, which was chosen to be the score threshold at the equal error operational point. This was proved to enhance the multi-biometric fusion performance, e.g. by reducing the FRR at 0.01%FAR by 44% to 75% in different experiment settings when moving from TanH normalization to the performance anchored PAN-TanH normalization.

The second pre-fusion topic dealt with missing data imputation in score-level multi-biometric fusion. In Section 3.4, a response to *RQ2* was drawn by stating the different performance responses to imputation methods between the verification and identification operational scenarios. For example, the minimum imputation rule achieved one of the highest identification performances while failing drastically in verification tasks. This pointed out the fact that a multi-biometric system should be specifically designed to fit its operational scenario. A more sophisticated solution was also proposed and tested in this work. These tests concluded that such a solution performs well under both scenarios. However, it does not significantly improve the multi-biometric performance in comparison to simpler imputation rules such as the mean rule.

Biometric source weighting is utilized to optimize the relative effect of the different biometric sources in combination-based score-level multi-biometric systems. This work focused on two aspects of assigning these weights. The first aspect is trying to define a weighting approach that captured both, the relative overall performance of each source and its relative confidence. This confidence is formulated as a measure of definition (not ambiguous) of the source decisions in error prone ranges. Such a weighting scheme was defined as a response to *RQ3* by combining a confidence measure and a performance measure represented by the EER. The confidence measure was defined as the standard deviation of the score distributions in the overlap area, where both genuine and imposter scores occur. The inclusion of this confidence measure through the proposed weight proved to outperform state-of-the-art and baseline solutions by achieving higher multi-biometric accuracy.

The second point of focus within multi-biometric source weighting was introduced by *RQ4* and is concerned with defining, and analyzing the effect, of a weighting scheme based on the relative identification performance of each source. Identification performance is concerned with a probe relation to a large number of references (1:N comparison). Therefore, a dynamic performance measures can be defined. Three main types of identification-based weights were formulated based on the cumulative identification rate at a certain rank, the rapidity in which the cumulative identification rate increase up to a certain rank, and the area above the CMC curve up to a certain rank. Weights defined by the identification rate at certain ranks performed best under the verification scenario, outperforming baseline solutions. Although performing well under the identification scenario, these weights were consistently outperformed in this scenario by the ones defined by the rate of improvement in cumulative identification rates.

As demonstrated in Figure 1.1, the fusion process can be supported by supplementary information to enhance the overall system accuracy and robustness. This information can be derived from raw captures at an early stage of the multi-biometric work-flow, which requires accessing and further processing of raw data. However, this work focused on extracting supplementary information that can be derived from the comparison scores. Two types of supplementary information are proposed and their effect on the system performance is analyzed. The first type is based on the relation between comparison scores between different comparisons. This was based on the assumption that a genuine score of a certain probe subject is relatively distanced from the set of a clustered set of imposter scores produced by the same probe. The ratios of the later distances to the first distance (most similar) is noted by neighbor distance ratios. These ratios were proposed as a response to *RQ5* and were proven to be discriminant between genuine and imposter comparisons. This type of supplementary information was integrated within a classification-based fusion approach. An improved solution was also presented by integrating biometric source weighting information in the NDR calculation process. This resulted in reducing the FRR by half at low FAR (0.001%) in most experiment settings.

The relation between different scores provided by a number of multi-biometric sources in one comparison was also discussed as a response to *RQ6*. Here, a hypothesis was made based on the assumption that the minority of biometric sources pointing out a different decision than the majority, might have faulty conclusions and should be given a relatively smaller role in the final fused decision. This was incorporated in a dynamic weighting approach that also considers static weights. This did not only prove to enhance the multi-biometric performance, but also maintain high accuracy when facing the more realistic scenario where some of the captured data could be slightly noisy. EER values under different test scenarios were reduced by at least 45% as a result of introducing coherence information.

Every biometric comparison requires access to the biometric reference database. Identification tasks, including duplicate enrollment checks, require a large number of comparisons that grows linearly with the size of the database. To enable a realistic response time for such tasks, database indexing approaches are utilized to limit the search range in the database, and thus limit the number of required comparisons. Multi-biometric provides the possibility of using richer information to achieve higher indexing performances, which leads to faster and more accurate searches. In this context, this work responded to *RQ7* by proposing a multi-biometric data retrieval approach that can combine single-source biometric indexing structures in a configurable and optimizable manner. This was performed by modifying the Borda count approach to be adaptable to sources of different performance nature, as well as allowing a user induced tuning on the fusion performance. This was also supplemented by the use of efficiently calculated approximated distances to enhance the retrieval performance while maintaining the flexibility of rank-level fusion approaches. Beside the general enhancement of the performance, this approach lead to a huge improvement in retrieval results in the critical cases where single sources supply low quality candidate lists.

In addition to processes in the core of the typical multi-biometric system, miscellaneous ones were also discussed in this work. These processes utilize and complement the core components of multi-biometric systems.

Three main points of focus were presented, the multi-biometric continuous authentication, face presentation attack detection, and creating unified face references from video sequences. In the aspect of multi-biometric continuous authentication, this work responded to *RQ8* by proposing an asynchronous fusion scheme that includes comparisons of keystroke dynamics and regularly captured face images. These comparisons contributed to a join trust value that can also trigger additional face captures. A realistic database was collected and an evaluation on this data showed that using the proposed fusion model can reduce false forced log-offs while maintaining high imposter detection rates. The design and implementation of the proposed solution concluded with a set of challenges and future work aspects that still face the deployment of such a system in realistic scenarios.

As an essential miscellaneous process in a multi-biometric system, this work also discussed face presentation attack detection. This was done by proposing a state-of-the-art solution based on analyzing the changes between video frames. The achieved results were compared with the reported performances in the literature. In a response to the *RQ9*, this work discussed realistic operational concerns. First by evaluating cross-database performance and concluding that most of the current solutions are sensitive to deployment conditions (different database) and defined this as an important focus point for future work. Feature and score-level fusion was also deployed and proved to enhance the detection performance. The score-level fusion over the video sequence showed that around 3 seconds of video capture is required to reach a saturated confident decision, and thus, processing longer captures are not beneficial.

The third miscellaneous process was concerned with creating an informative single face representation template from a video sequence. This is essential to minimize the processing effort needed in unconstrained face recognition from videos. To achieve this, and as a response to *RQ10*, two main aspects where tackled. First is the selection of key representative faces from the video sequence to use in creating the fused template. Here, different approaches were evaluated with face selected by detection confidence and by being most different achieved the best results. The second aspect is the feature fusion approach, where the evaluation showed the superiority of voting and discriminate-voting approaches. The proposed key-face selection and feature-level fusion approaches proved to, at least, match the performance of the computationally intensive score-level fusion of all possible face comparisons between two video sequences.

The previously discussed contributions are part of the different components of the typical multi-biometric system discussed in Chapter 1. Figure 8.1 links the multi-biometric work-flow to the research questions introduces in Section 1.2, the chapters that provide a response to these questions, the publications on which these chapters were based, and a hint on the proposed solutions.

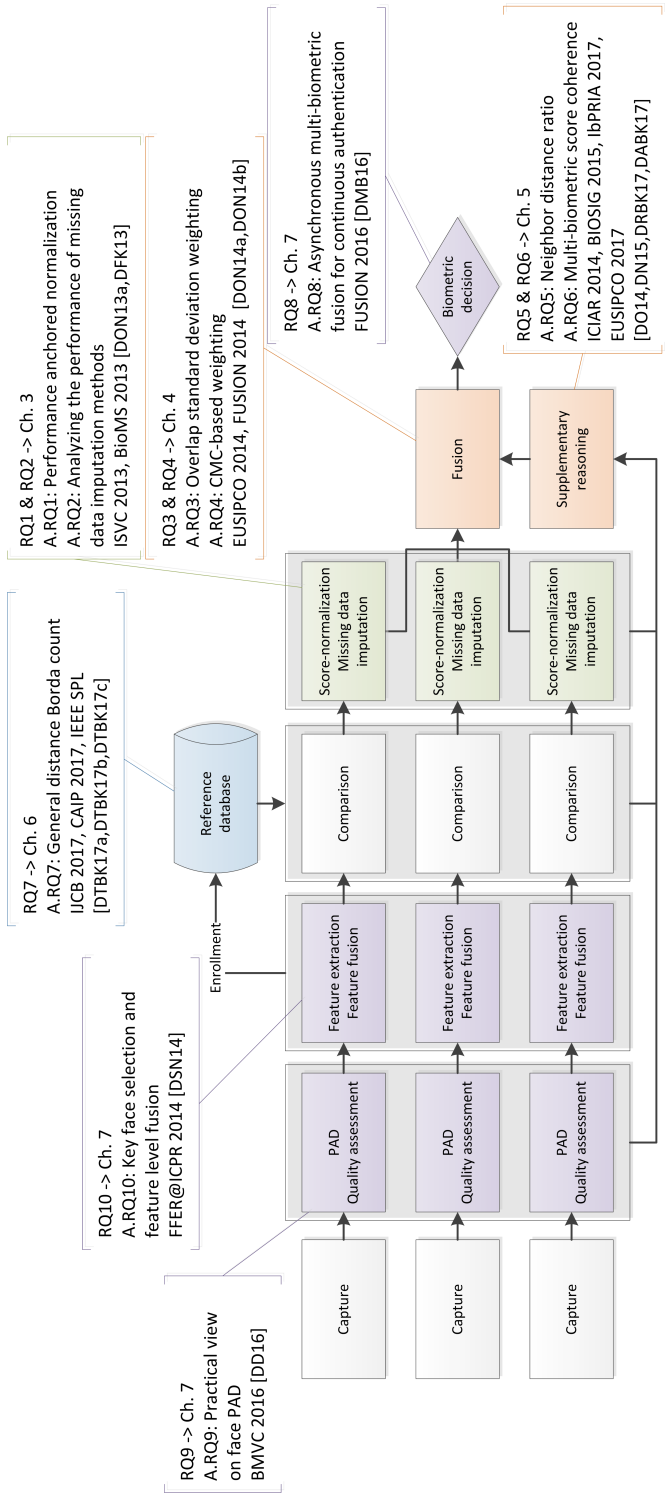


Figure 8.1: the work-flow of a multi-biometric system tagged with the research questions, the chapters responding to these questions, the publications building up to these chapters, and a hint at the proposed solutions.

8.2 Future work

As multi-biometrics is enabling both large-scale identity management and user centric biometric systems, these systems are ever more demanding, and thus pushing future research, in terms of performance, usability, and security.

Re-identification biometric-based surveillance and systems that aim at identifying black listed individuals on the fly usually use the same back-end biometric comparison algorithms as the ones designed for more conventional verification/identification systems, where the subjects intend to be verified/identified. The conventional algorithms assume that the subject or operator intends to recognize the identity and thus focuses on preventing attacks that try to make the attacker appear as a different enrolled subject. Re-identification systems face a different kind of attacks where an enrolled subject tries to appear as an imposter to the references in the database. To prevent such attacks, these systems should be designed and optimized differently. A specially designed multi-biometric solution can provide a boost of protection to such systems, especially with the increased difficulty of performing presentation attacks on multiple modalities/algorithms. The design, optimization, and evaluation of such a system is yet to be studied.

This work included unconventional information in the fusion process in the form of neighbor distance ratios and score coherence. Further available information can be probed as valuable additions to increase the biometric decision accuracy. Such information can result from the definition of the biometric *Doddington zoo* [DLM*98], where individuals and their biometric characteristics are divided into groups based on their relation to each other. These categories have been shown to be statistically valid and to have an effect on the biometric comparison. However, due to the limited size of biometric databases and the absence of a suitable representation, an automatic classification of subjects into these categories is still an open issue. Building a well performing automatic classification solution can open the door to include such information in the fusion process and improving the performance stability of multi-biometric systems.

Deep learning techniques are taking the artificial intelligence industry by the storm. Biometric solutions have recently gained huge accuracy momentum by using deep learning techniques along with large databases, especially for face recognition. Using deep learning to create discriminant joint representations from multiple biometric modalities is an interesting case to study, including investigating the optimal point in the network to join the modalities. Besides the expected highly discriminant representation of the identity, this representation will be inherently more secure against attacks as all the modalities involved have to be presented to create a valid representation. The joint presentation also enables new deployment strategies for multi-biometric template protection.

Despite numerous effort to make biometric systems more secure, especially when compared to conventional proofs of identity, attackers still find innovative solution to both circumvent biometric authentication or retrieve stored biometric data. *Protecting stored biometric templates* is a well studied topic. However, taking advantage of the properties multi-biometric data to increase the security of stored templates, including unified multi-biometric templates, is a more recent topic and a very promising research field. Another point of attack on biometric systems is by directly presenting counterfeited biometric characteristics to the capturing sensor, i.e. presentation attack. Solution for detecting presentation attacks for different individual modalities have been studied recently in the literature and took advantage of multi-biometric fusion techniques, including this work. Multi-modal biometric systems are inherently more difficult to attack, as they need multiple presentation attacks to succeed. Creating, analyzing, and evaluating multi-biometric attacks and *multi-biometric presentation attack detection* solutions is a promising and necessary task to perform in the future.

Biometric solutions deployed on consumer electronics are gaining popularity, especially in the financial applications. Different device manufacturers and models include variations of sensors and signal processing im-

plementations. This results in degrading the *biometric cross-device performance*. Some solutions have been proposed to utilize multi-biometric fusion to enhance cross-device verification performance. However, with the unrealistic assumption of having prior knowledge about all devices [ARBB17]. Advanced fusion approaches can be utilized to adapt to different devices without any prior knowledge and maintain acceptable performance stability. This will allow more accurate biometric verification on consumer electronics, which will enable new applications.

A Publications and talks

The thesis is partially based on the following publications and talks:

A.1 Publications

1. DAMER N., TERHÖRST P., BRAUN A., KUIJPER A.: General borda count for multi-biometric retrieval. In *IEEE International Joint Conference on Biometrics, IJCB 2017, Denver, Colorado, USA, October 1 - 4, 2017* (2017), IEEE
2. DAMER N., RHAIBANI C. I., BRAUN A., KUIJPER A.: Trust the biometrie mainstream: Multi-biometric fusion and score coherence. In *25th European Signal Processing Conference, EUSIPCO 2017, Kos, Greece, August 28 - September 2, 2017* (2017), IEEE, pp. 2191–2195
3. DAMER N., TERHÖRST P., BRAUN A., KUIJPER A.: Efficient, accurate, and rotation-invariant iris code. *IEEE Signal Processing Letters 24*, 8 (Aug 2017), 1233–1237
4. DAMER N., TERHÖRST P., BRAUN A., KUIJPER A.: Indexing of single and multi-instance iris data based on lsh-forest and rotation invariant representation. In *Computer Analysis of Images and Patterns - 17th International Conference, CAIP 2017, Ystad, Sweden, August 22-24, 2017, Proceedings, Part II* (2017), Felsberg M., Heyden A., Krüger N., (Eds.), vol. 10425 of *Lecture Notes in Computer Science*, Springer, pp. 190–201
5. DAMER N., ALKHATIB W., BRAUN A., KUIJPER A.: Neighbor distance ratios and dynamic weighting in multi-biometric fusion. In *Pattern Recognition and Image Analysis - 8th Iberian Conference, IbPRIA 2017, Faro, Portugal, June 20-23, 2017, Proceedings* (2017), Alexandre L. A., Sánchez J. S., Rodrigues J. M. F., (Eds.), vol. 10255 of *Lecture Notes in Computer Science*, Springer, pp. 491–500
6. DAMER N., DIMITROV K.: Practical view on face presentation attack detection. In *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016* (2016), Wilson R. C., Hancock E. R., Smith W. A. P., (Eds.), BMVA Press
7. DAMER N., MAUL F., BUSCH C.: Multi-biometric continuous authentication: A trust model for an asynchronous system. In *2016 19th International Conference on Information Fusion (FUSION)* (July 2016), pp. 2192–2199
8. DAMER N., NOUAK A.: Weighted integration of neighbors distance ratio in multi-biometric fusion. In *BIOSIG 2015 - Proceedings of the 14th International Conference of the Biometrics Special Interest Group, 9.-11. September 2015, Darmstadt, Germany* (2015), Brömme A., Busch C., Rathgeb C., Uhl A., (Eds.), vol. 245 of *LNI*, GI, pp. 255–262
9. DAMER N., OPEL A.: Multi-biometric score-level fusion and the integration of the neighbors distance ratio. In *11th International Conference on Image Analysis and Recognition, ICIAR 2014, Vilamoura, Portugal, October 22-24, 2014, Proceedings, Part II* (2014), Campilho A. J. C., Kamel M. S., (Eds.), vol. 8815 of *Lecture Notes in Computer Science*, Springer, pp. 85–93

10. DAMER N., OPEL A., NOUAK A.: CMC curve properties and biometric source weighting in multi-biometric score-level fusion. In *17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, July 7-10, 2014* (2014), IEEE, pp. 1–6

11. DAMER N., OPEL A., NOUAK A.: Biometric source weighting in multi-biometric fusion: Towards a generalized and robust solution. In *22nd European Signal Processing Conference, EUSIPCO 2014, Lisbon, Portugal, September 1-5, 2014* (2014), IEEE, pp. 1382–1386

12. DAMER N., SAMARTZIDIS T., NOUAK A.: Personalized face reference from video: Key-face selection and feature-level fusion. In *Face and Facial Expression Recognition from Real World Videos - International Workshop, FFER@ICPR 2014, Stockholm, Sweden, August 24, 2014, Revised Selected Papers* (2014), Ji Q., Moeslund T. B., Hua G., Nasrollahi K., (Eds.), vol. 8912 of *Lecture Notes in Computer Science*, Springer, pp. 85–98

13. DAMER N., OPEL A., SHAHVERDYAN A.: An overview on multi-biometric score-level fusion - verification and identification. In *ICPRAM 2013 - Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods, Barcelona, Spain, 15-18 February, 2013.* (2013), Marsico M. D., Fred A. L. N., (Eds.), SciTePress, pp. 647–653

14. DAMER N., FÜHRER B., KUIJPER A.: Missing data estimation in multi-biometric identification and verification. In *2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications* (Sept 2013), pp. 41–45

15. DAMER N., OPEL A., NOUAK A.: Performance anchored score normalization for multi-biometric fusion. In *9th International Symposium on Advances in Visual Computing, ISVC 2013, Rethymnon, Crete, Greece, July 29-31, 2013. Proceedings, Part II* (2013), Bebis G., Boyle R., Parvin B., Koracin D., Li B., Porikli F., Zordan V. B., Klosowski J. T., Coquillart S., Luo X., Chen M., Gotz D., (Eds.), vol. 8034 of *Lecture Notes in Computer Science*, Springer, pp. 68–75

16. DAMER N., FÜHRER B.: Ear recognition using multi-scale histogram of oriented gradients. In *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2012, Piraeus-Athens, Greece, July 18-20, 2012* (2012), Tsihrintzis G. A., Pan J., Huang H., Virvou M., Jain L. C., (Eds.), IEEE, pp. 21–24

17. SAMARTZIDIS T., SIEGMUND D., GOEDDE M., DAMER N., BRAUN A., KUIJPER A.: The dark side of the face: exploring the ultraviolet spectrum for face biometrics. In *International Conference on Biometrics, ICB 2018, 20-23 February, 2018, Gold Coast, Queensland, Australia* (2018), IEEE (to appear)

18. HENNIGER O., DAMER N., BRAUN A.: Opportunities for biometric technologies in smart environments. In *Ambient Intelligence - 13th European Conference, AmI 2017, Malaga, Spain, April 26-28, 2017, Proceedings* (2017), Braun A., Wichert R., Maña A., (Eds.), vol. 10217 of *Lecture Notes in Computer Science*, pp. 175–182

19. SIEGMUND D., EBERT T., DAMER N.: Combining low-level features of offline questionnaires for handwriting identification. In *Image Analysis and Recognition - 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel, Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings* (2016), Campilho A., Karray F., (Eds.), vol. 9730 of *Lecture Notes in Computer Science*, Springer, pp. 46–54

20. MAUL F., DAMER N.: Fuzzy logic and multi-biometric fusion - an overview. In *ICPRAM 2015 - Proceedings of the International Conference on Pattern Recognition Applications and Methods, Volume 1, Lisbon, Portugal, 10-12 January, 2015.* (2015), Marsico M. D., Figueiredo M. A. T., Fred A. L. N., (Eds.), SciTePress, pp. 218–222

21. BUTT M., DAMER N., RATHGEB C.: Privacy preserved duplicate check using multi-biometric fusion. In *17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, July 7-10, 2014* (2014), IEEE, pp. 1–7
22. SIEGMUND D., SAMARTZIDIS T., DAMER N., NOUAK A., BUSCH C.: Virtual fitting pipeline: Body dimension recognition, cloth modeling, and on-body simulation. In *VRIPHYS 14: 11th Workshop on Virtual Reality Interactions and Physical Simulations, Bremen, Germany, 2014. Proceedings* (2014), Bender J., Duriez C., Jaillet F., Zachmann G., (Eds.), Eurographics Association, pp. 99–107
23. BUTT M., DAMER N.: Helper data scheme for 2d cancelable face recognition using bloom filters. In *2014 International Conference on Systems, Signals and Image Processing, IWSSIP 2014* (May 2014), pp. 271–274
24. IRUJO J. A., CUADROS M., AGINAKO N., RAFFAELLI M., KÄHM O., DAMER N., NETO J. P.: Multimedia analysis of video sources. In *SIGMAP 2014 - Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications, Vienna, Austria, 28-30 August, 2014* (2014), Obaidat M. S., Holzinger A., Cabello E., (Eds.), SciTePress, pp. 346–352
25. CHINGOVSKA I., YANG J., LEI Z., YI D., LI S. Z., KÄHM O., GLASER C., DAMER N., KUIJPER A., NOUAK A., KOMULAINEN J., PEREIRA T. F., GUPTA S., KHANDELWAL S., BANSAL S., RAI A., KRISHNA T., GOYAL D., WARIS M., ZHANG H., AHMAD I., KIRANYAZ S., GABBOUJ M., TRONCI R., PILI M., SIRENA N., ROLI F., GALBALLY J., FIÉRREZ J., DA SILVA PINTO A., PEDRINI H., SCHWARTZ W. S., ROCHA A., ANJOS A., MARCEL S.: The 2nd competition on counter measures to 2d face spoofing attacks. In *International Conference on Biometrics, ICB 2013, 4-7 June, 2013, Madrid, Spain* (2013), Fiérrez J., Kumar A., Vatsa M., Veldhuis R. N. J., Ortega-Garcia J., (Eds.), IEEE, pp. 1–6
26. KÄHM O., DAMER N.: 2d face liveness detection: An overview. In *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group, Darmstadt, Germany, September 6-7, 2012* (2012), Brömme A., Busch C., (Eds.), vol. 197 of *LNI*, IEEE/GI, pp. 1–12

For an updated publication list, please check my DBLP and Google Scholar profiles:

- DBLP: http://dblp.uni-trier.de/pers/hd/d/Damer:Naser
- Google Scholar: https://scholar.google.de/citations?user=bAyT17sAAAAJ&hl=en

A.2 Talks

1. Damer N., Bringer J.: Strengthening fight against identity fraud at enrolment: use of multiple biometrics and privacy perspectives. European Conference on ePassports, Brussels, Belgium, 10 December 2015
2. Damer N.: Multi-biometric fusion. da/sec Scientifc Talk, Darmstadt, Germany, 28 January 2013
3. Braun A., Damer N.: Detecting attacks on biometric recognition systems . Wissenschaftstag spezial: "Cybersicherheitsforschung aus Darmstadt - Sicherheit und Datenschutz im Internet" , Darmstadt, Germany, 10 September 2017

A.3 Posters

1. Damer N., Dimitrov K.: Practical View on Face Presentation Attack Detection. Besuch der AG Forschung, Vertretung der Ministerien HMWK, HMWVL, Hessische Staatskanzlei, HTAI, Hessen Agentur , TU Darmstadt, Darmstadt, Germany, 13 December 2016
2. Damer N.: Score-level Multi-biometric Fusion for Duplicate Enrollment Check. FIDELITY Conference, Brussels, Belgium, 10-11 December 2015
3. Mansfield T., Shenoy A., Henniger O., Damer N., Gacon P.: Case Study: Evaluation & Certification of Biometric Products. HECTOS Final Event, BAO Congress Centre, Brussels, Belgium, 6 December 2017

A.4 Submitted papers

1. DAMER N., BOUTROS F., TERHÖRST P., BRAUN A., KUIJPER A.: P-score: Performance aligned normalization and an evaluation in score-level multi-biometric fusion. In *26th European Signal Processing Conference, EUSIPCO 2018, Rome, Italy, September 3 - 7, 2018* (2018), IEEE (under review)
2. DAMER N., DIMITROV K., BRAUN A., KUIJPER A.: On creating joint multi-biometric representations by deep fusion. In *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 2018* (2018), IEEE (under review)
3. DAMER N., WAINAKH Y., HENNIGER O., CROLL C., BERTHE B., BRAUN A., KUIJPER A.: Deep learning-based face recognition and the robustness to perspective distortion. In *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 2018* (2018), IEEE (under review)
4. TERHÖRST P., DAMER N., BOUTROS F., BRAUN A., KUIJPER A.: What can a single minutia tell about gender? In *6th IAPR/IEEE International Workshop on Biometrics and Forensics, Sassari, IT, June 7, 8 2018* (2018), IEEE (under review)

B Supervising activities

The following list summarizes the student bachelor, diploma and master thesis supervised by the author. The results of these works were partially used as an input into the thesis.

B.1 Diploma and master thesis

1. Benedikt Führer, Arjan Kuijper (supervisor), and Naser Damer (supervisor). Quality-based score-level fusion for multi-biometrics under identification framework. TU Darmstadt, Master Thesis, 2013
2. Christian Glaser, Arjan Kuijper (supervisor), Olga Kähm (supervisor), and Naser Damer (supervisor). Face liveness detection against image and video spoofing attacks. TU Darmstadt, Master Thesis, 2013
3. Dirk Siegmund, Christoph Busch (supervisor), and Naser Damer (supervisor). Body dimension detection for an in-shop advertisement system. Hochschule Darmstadt, Master Thesis, 2014
4. Timotheos Samartzidis, Christoph Busch (supervisor), and Naser Damer (supervisor). Automated clothes 3D reconstruction for an in-shop advertisement system. Hochschule Darmstadt, Master Thesis, 2014
5. Fabian Maul, Christoph Busch (supervisor), and Naser Damer (supervisor). Multi-biometric continuous authentication. Hochschule Darmstadt, Master Thesis, 2015
6. Philipp Terhörst, Thomas Walther (supervisor), and Naser Damer (supervisor). Indexing of multi-biometrics databases: fast and accurate biometric search. TU Darmstadt, Master Thesis, 2017 (IGD Best Thesis Award)
7. Kristiyan Dimitrov, Arjan Kuijper (supervisor), and Naser Damer (supervisor). Exploring deep multi-biometric fusion. TU Darmstadt, Master Thesis, 2017 (IGD Best Thesis Award Nominee)
8. Roshan N. Anvekar, Andreas Pech (supervisor), Naser Damer (supervisor). Cross-device biometric verification and the benefit of multi-algorithmic biometric fusion. Frankfurt UAS, Master Thesis, 2018 (ongoing)
9. Meltem Subasioglu, Naser Damer (supervisor). On predicting people professions from their face images: human vs. machine. Goethe Universität Frankfurt, Master Thesis, 2018 (ongoing)
10. Alexandra Mosegui, Alain Tremeau (supervisor), and Naser Damer (supervisor). Novel face morphing attacks and the vulnerabilities of the morphing attack detectors. Jean Monnet University Saint-Etienne, Master Thesis, 2018 (ongoing)

B.2 Bachelor thesis

1. Fabian Maul, Christoph Busch (supervisor), and Naser Damer (supervisor). Analysis of normalization and combination algorithms in the context of multi-biometric score-level fusion. Hochschule Darmstadt, Bachelor Thesis, 2013

2. Marianne Melzer, Karin Schwarz (supervisor), and Naser Damer (supervisor). Face recognition software in image archives - An investigation into the use of facial recognition programs for the archaic core tasks "review" and "development". Fachhochschule Potsdam, Bachelor Thesis, 2014
3. Anumudu Lawrence Kelechi, Marian Margraf (supervisor), and Naser Damer (supervisor). Utilizing fuzzy distance in biometric comparison. Hochschule Darmstadt, Bachelor Thesis, 2015
4. Sven von den Berken, Ralf Hahn (supervisor), Naser Damer (supervisor). Face presentation attacks on identification systems. Hochschule Darmstadt, Bachelor Thesis, 2018 (ongoing)

C Curriculum vitae

Personal data

Name	Naser Damer
Birth date & place	14.2.1985 in Amman, Jordan

Education

2008-2010	Master of Science in Electrical and Computer Engineering (Automation and Control), Technical University of Kaiserslautern, Kaiserslautern, Germany
2003-2008	Bachelor of Science in Mechatronics Engineering, University of Jordan, Amman, Jordan

Work experience

Oct 2011 – current Researcher, Fraunhofer Institute for Computer Graphics Research IGD, Darmstadt, Germany

Selected projects

2011 – 2015	GES-3D: Multi-biometric person identification
2011 – 2014	CAPER : Collaborative information, Acquisition, Processing, Exploitation and Reporting for the prevention of organized crime
2012 – 2015	FIDELITY: Fast and trustworthy Identity Delivery and check with ePassports leveraging Traveler privacy
2014 – present	HECTOS: Harmonized Evaluation, Certification and Testing of Security Products
2015 – present	YAQEEN ABIS: Providing technical and scientific consultancy to the Royal Oman Police in creating a large scale modern automated biometric identification system (ABIS)
2016 – present	CRISP: Center for Research in Security and Privacy

Mar 2010–Sep 2010 Master thesis work, Robert Bosch GmbH, Chassis Systems Control, Leonberg, Germany

Dec 2009–Mar 2010 Internship, Robert Bosch GmbH, Chassis Systems Control, Leonberg, Germany

Jun 2007–Aug 2007 Internship, Industrija Kablova – Jagodina (FKS), Jagodina, Serbia

Academic activities

Organizer	Organizing and chairing the special session Information Fusion in Multi-Biometrics and Forensics at the International Conference on Information FUSION 2013-2018
PC member	Technical program committee member at the International Conference on Information FUSION 2013-2018
Review	Reviewer for a number of journals, conferences, and awards: IEEE signal processing magazine, IEEE Access, Multimedia Tools and Applications, IJCB, BIOSIG, FUSION, WIBC, CAST e.V. den Förderpreis IT-Sicherheit award.
Honors	Biometric summer school scholarship supported by COST Action IC1106
	Selected to participate in the global young scientists summit@one-north 2015, Singapore
Teaching	Security in ambient intelligent, "ambient intelligent (AmI)", TU Darmstadt, 2017
	Supervisor for a number of practical course works within "Programming a Graphic System", GRIS, TU Darmstadt 2012-2017

Bibliography

[AH06] ARANDJELOVIC O., HAMMOUD R.: Multi-sensory face biometric fusion (for personal identification). In *Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop* (Washington, DC, USA, 2006), CVPRW '06, IEEE Computer Society, pp. 128–. 15

[Als10] ALSAADE F.: A study of neural network and its properties of training and adaptability in enhancing accuracy in a multimodal biometrics scenario. *Information Technology Journal 9* (January 2010), 188–191. 17, 56, 89

[ALS16] AMOS B., LUDWICZUK B., SATYANARAYANAN M.: *OpenFace: A general-purpose face recognition library with mobile applications*. Tech. rep., CMU-CS-16-118, CMU School of Computer Science, 2016. 65

[AM11] ANJOS A., MARCEL S.: Counter-measures to photo attacks in face recognition: A public database and a baseline. In *2011 IEEE International Joint Conference on Biometrics, IJCB 2011, Washington, DC, USA, October 11-13, 2011* (2011), IEEE, pp. 1–7. 99

[AMSS08] AZZINI A., MARRARA S., SASSI R., SCOTTI F.: A fuzzy approach to multimodal biometric continuous authentication. *Fuzzy Optimization and Decision Making 7*, 3 (2008), 243–256. 88

[APM*12] AFONSO L. C. S., PAPA J. P., MARANA A. N., POURSABERI A., YANUSHKEVICH S. N.: A fast large scale iris database classification with optimum-path forest technique: A case study. In *The 2012 International Joint Conference on Neural Networks (IJCNN), Brisbane, Australia, June 10-15, 2012* (2012), pp. 1–5. 71

[ARBB17] ALONSO-FERNANDEZ F., RAJA K. B., BUSCH C., BIGÜN J.: Log-likelihood score level fusion for improved cross-sensor smartphone periocular recognition. In *25th European Signal Processing Conference, EUSIPCO 2017, Kos, Greece, August 28 - September 2, 2017* (2017), IEEE, pp. 271–275. 122

[BB09] BOURS P., BARGHOUTHI H.: Continuous authentication using biometric keystroke dynamics. In *The Norwegian Information Security Conference (NISK)* (2009), vol. 2009. 85, 86, 89, 90, 93

[BBC02] BELLOT D., BOYER A., CHARPILLET F.: A new definition of qualified gain in a data fusion process: application to telemedicine. In *Proceedings of the Fifth International Conference on Information Fusion. FUSION 2002.* (July 2002), vol. 2, pp. 865–872 vol.2. 13

[BBKQ09] BAIG A., BOURIDANE A., KURUGOLLU F., QU G.: Fingerprint - iris fusion based identification system using a single hamming distance matcher. In *Proceedings of the 2009 Symposium on Bio-inspired Learning and Intelligent Systems for Security* (Washington, DC, USA, 2009), BLISS '09, IEEE Computer Society, pp. 9–12. 14

[BCG05] BAWA M., CONDIE T., GANESAN P.: Lsh forest: Self-tuning indexes for similarity search. In *Proceedings of the 14th International Conference on World Wide Web* (New York, NY, USA, 2005), WWW '05, ACM, pp. 651–660. 74

[BD14] BUTT M., DAMER N.: Helper data scheme for 2d cancelable face recognition using bloom filters. In *2014 International Conference on Systems, Signals and Image Processing, IWSSIP 2014* (May 2014), pp. 271–274. 125

[BDR14] BUTT M., DAMER N., RATHGEB C.: Privacy preserved duplicate check using multi-biometric fusion. In *17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, July 7-10, 2014* (2014), IEEE, pp. 1–7. 125

[BDVS13] BHARADWAJ S., DHAMECHA T. I., VATSA M., SINGH R.: Computationally efficient face spoofing detection with motion magnification. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops* (June 2013), pp. 105–110. 99, 103, 105

[BHK97] BELHUMEUR P. N., HESPANHA J. P., KRIEGMAN D. J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence 19*, 7 (1997), 711–720. 88, 106, 108

[BKTR10] BASAK J., KATE K., TYAGI V., RATHA N. K.: A gradient descent approach for multi-modal biometric identification. In *20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010* (2010), IEEE Computer Society, pp. 1322–1325. 15, 32

[BLLJ09] BAO W., LI H., LI N., JIANG W.: A liveness detection method for face recognition based on optical flow field. In *2009 International Conference on Image Analysis and Signal Processing* (April 2009), pp. 233–236. 99

[BM15] BOURS P., MONDAL S.: *Recent Advances in User Authentication Keystroke Dynamics Biometrics*. Science Gate Publishing, 2015, ch. Continuous Authentication with Keystroke Dynamics, pp. 41–58. 7, 85

[Bou12] BOURS P.: Continuous keystroke dynamics: A different perspective towards biometric evaluation. *Information Security Technical Report 17*, 1 (2012), 36–43. 85, 89

[BW12] BANERJEE S. P., WOODARD D. L.: Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research 7*, 1 (2012), 116–139. 7, 85, 88, 89

[CAM12] CHINGOVSKA I., ANJOS A., MARCEL S.: On the effectiveness of local binary patterns in face anti-spoofing. In *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group, Darmstadt, Germany, September 6-7, 2012* (2012), Brömme A., Busch C., (Eds.), vol. 197 of *LNI*, IEEE/GI, pp. 1–7. 86, 99, 105

[CBFC04] CHANG K. I., BOWYER K. W., FLYNN P. J., CHEN X.: Multi-biometrics using facial appearance, shape and temperature. In *Proceedings of the Sixth IEEE international conference on Automatic face and gesture recognition* (Washington, DC, USA, 2004), FGR' 04, IEEE Computer Society, pp. 43–48. 17, 26

[CBSV03] CHANG K., BOWYER K. W., SARKAR S., VICTOR B.: Comparison and combination of ear and face images in appearance-based biometrics. *IEEE Trans. Pattern Analysis and Machine Intelligence 25* (2003), 1160–1165. 14

[CFM10] CAPPELLI R., FERRARA M., MALTONI D.: Minutia cylinder-code: A new representation and matching technique for fingerprint recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence 32* (2010), 2128–2141. 77

[CFMT10] CAPPELLI R., FERRARA M., MALTONI D., TISTARELLI M.: MCC: A baseline algorithm for fingerprint verification in fvc-ongoing. In *11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010, Singapore, 7-10 December 2010, Proceedings* (2010), IEEE, pp. 19–23. 77

[Chi13] CHINESE ACADEMY OF SCIENCES, INSTITUTE OF AUTOMATION:. `http://english.ia.cas.cn/`, 2013. 77

[CLX*13] CUI Z., LI W., XU D., SHAN S., CHEN X.: Fusing robust face region descriptors via multiple metric learning for face recognition in the wild. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013* (2013), IEEE Computer Society, pp. 3554–3561. 107, 112

[CMM00] CAPPELLI R., MAIO D., MALTONI D.: Combining fingerprint classifiers. In *Proceedings of the First International Workshop on Multiple Classifier Systems* (London, UK, UK, 2000), MCS '00, Springer-Verlag, pp. 351–361. 16, 25

[CMM02] CAPPELLI R., MAIO D., MALTONI D.: Synthetic fingerprint-database generation. In *Proceedings of the 16 th International Conference on Pattern Recognition (ICPR'02) Volume 3 - Volume 3* (Washington, DC, USA, 2002), ICPR '02, IEEE Computer Society, pp. 30744–. 77

[CotEU04] COUNCIL OF THE EUROPEAN UNION .: *2004/512/EC: Council Decision of 8 June 2004 establishing the Visa Information System (VIS)*. Decision, Council of the European Union, June 2004. 1

[CRHV09] CHAUDHRY R., RAVICHANDRAN A., HAGER G. D., VIDAL R.: Histograms of oriented optical flow and binet-cauchy kernels on nonlinear dynamical systems for the recognition of human actions. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA* (2009), IEEE Computer Society, pp. 1932–1939. 99

[CSN10] CHIA C., SHERKAT N., NOLLE L.: Towards a best linear combination for multimodal biometric fusion. In *20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010* (2010), IEEE Computer Society, pp. 1176–1179. 38, 39, 56, 57, 89, 107

[CTG11] CHENG X., TULYAKOV S., GOVINDARAJU V.: Multiple-sample fusion of matching scores in biometric systems. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2011, Colorado Springs, CO, USA, 20-25 June, 2011* (2011), IEEE Computer Society, pp. 120–125. 15

[CYL*13] CHINGOVSKA I., YANG J., LEI Z., YI D., LI S. Z., KÄHM O., GLASER C., DAMER N., KUIJPER A., NOUAK A., KOMULAINEN J., PEREIRA T. F., GUPTA S., KHANDELWAL S., BANSAL S., RAI A., KRISHNA T., GOYAL D., WARIS M., ZHANG H., AHMAD I., KIRANYAZ S., GABBOUJ M., TRONCI R., PILI M., SIRENA N., ROLI F., GALBALLY J., FIÉRREZ J., DA SILVA PINTO A., PEDRINI H., SCHWARTZ W. S., ROCHA A., ANJOS A., MARCEL S.: The 2nd competition on counter measures to 2d face spoofing attacks. In *International Conference on Biometrics, ICB 2013, 4-7 June, 2013, Madrid, Spain* (2013), Fiérrez J., Kumar A., Vatsa M., Veldhuis R. N. J., Ortega-Garcia J., (Eds.), IEEE, pp. 1–6. 125

[DABK17] DAMER N., ALKHATIB W., BRAUN A., KUIJPER A.: Neighbor distance ratios and dynamic weighting in multi-biometric fusion. In *Pattern Recognition and Image Analysis - 8th Iberian Conference, IbPRIA 2017, Faro, Portugal, June 20-23, 2017, Proceedings* (2017), Alexandre L. A., Sánchez J. S., Rodrigues J. M. F., (Eds.), vol. 10255 of *Lecture Notes in Computer Science*, Springer, pp. 491–500. 55, 123

[DAGG11] DHALL A., ASTHANA A., GOECKE R., GEDEON T.: Emotion recognition using PHOG and LPQ features. In *Ninth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011), Santa Barbara, CA, USA, 21-25 March 2011* (2011), IEEE Computer Society, pp. 878–883. 87

[Dau04] DAUGMAN J.: How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology 14*, 1 (Jan 2004), 21–30. 72

[DBK*96] DRUCKER H., BURGES C. J. C., KAUFMAN L., SMOLA A. J., VAPNIK V.: Support vector regression machines. In *Advances in Neural Information Processing Systems 9, NIPS, Denver, CO, USA, December 2-5, 1996* (1996), Mozer M., Jordan M. I., Petsche T., (Eds.), MIT Press, pp. 155–161. 31

[DBT*18] DAMER N., BOUTROS F., TERHÖRST P., BRAUN A., KUIJPER A.: P-score: Performance aligned normalization and an evaluation in score-level multi-biometric fusion. In *26th European Signal Processing Conference, EUSIPCO 2018, Rome, Italy, September 3 - 7, 2018* (2018), IEEE (under review). 126

[DD16] DAMER N., DIMITROV K.: Practical view on face presentation attack detection. In *Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19-22, 2016* (2016), Wilson R. C., Hancock E. R., Smith W. A. P., (Eds.), BMVA Press. 85, 123

[DDBK18] DAMER N., DIMITROV K., BRAUN A., KUIJPER A.: On creating joint multi-biometric representations by deep fusion. In *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 2018* (2018), IEEE (under review). 126

[DDV07] DINERSTEIN S., DINERSTEIN J., VENTURA D.: Robust multi-modal biometric fusion via multiple svms. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, MontrÃl'al, Canada, 7-10 October 2007* (2007), IEEE, pp. 1530–1535. 17

[DF12] DAMER N., FÜHRER B.: Ear recognition using multi-scale histogram of oriented gradients. In *Eighth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP 2012, Piraeus-Athens, Greece, July 18-20, 2012* (2012), Tsihrintzis G. A., Pan J., Huang H., Virvou M., Jain L. C., (Eds.), IEEE, pp. 21–24. 124

[DFK13] DAMER N., FÜHRER B., KUIJPER A.: Missing data estimation in multi-biometric identification and verification. In *2013 IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications* (Sept 2013), pp. 41–45. 21, 124

[dFPAMM12] DE FREITAS PEREIRA T., ANJOS A., MARTINO J. M. D., MARCEL S.: LBP - TOP based countermeasure against face spoofing attacks. In *Computer Vision - ACCV 2012 Workshops - ACCV 2012 International Workshops, Daejeon, Korea, November 5-6, 2012, Revised Selected Papers, Part I* (2012), Park J., Kim J., (Eds.), vol. 7728 of *Lecture Notes in Computer Science*, Springer, pp. 121–132. 105

[dFPAMM13] DE FREITAS PEREIRA T., ANJOS A., MARTINO J. M. D., MARCEL S.: Can face anti-spoofing countermeasures work in a real world scenario? In *2013 International Conference on Biometrics (ICB)* (June 2013), pp. 1–8. 105

[DLM*98] DODDINGTON G. R., LIGGETT W., MARTIN A. F., PRZYBOCKI M. A., REYNOLDS D. A.: Sheep, goats, LAMBS and WOLVES: a statistical analysis of speaker performance in the NIST 1998 speaker recognition evaluation. In *The 5th International Conference on Spoken Language Processing, Incorporating The 7th Australian International Speech Science and Technology Conference, Sydney Convention Centre, Sydney, Australia, 30th November - 4th December 1998* (1998), ISCA. 121

[DMB16] DAMER N., MAUL F., BUSCH C.: Multi-biometric continuous authentication: A trust model for an asynchronous system. In *2016 19th International Conference on Information Fusion (FUSION)* (July 2016), pp. 2192–2199. 85, 123

[DN15] DAMER N., NOUAK A.: Weighted integration of neighbors distance ratio in multi-biometric fusion. In *BIOSIG 2015 - Proceedings of the 14th International Conference of the Biometrics Special Interest Group, 9.-11. September 2015, Darmstadt, Germany* (2015), Brömme A., Busch

C., Rathgeb C., Uhl A., (Eds.), vol. 245 of *LNI*, GI, pp. 255–262. 55, 123

[DNN13] DEUTSCHMANN I., NORDSTROM P., NILSSON L.: Continuous authentication using behavioral biometrics. *IT Professional 15*, 4 (2013), 12–15. 85

[DNRT11] DE MARSICO M., NAPPI M., RICCIO D., TORTORA G.: NABS: novel approaches for biometric systems. *IEEE Trans. Systems, Man, and Cybernetics, Part C 41*, 4 (2011), 481–493. 16

[DO14] DAMER N., OPEL A.: Multi-biometric score-level fusion and the integration of the neighbors distance ratio. In *11th International Conference on Image Analysis and Recognition, ICIAR 2014, Vilamoura, Portugal, October 22-24, 2014, Proceedings, Part II* (2014), Campilho A. J. C., Kamel M. S., (Eds.), vol. 8815 of *Lecture Notes in Computer Science*, Springer, pp. 85–93. 55, 56, 89, 123

[DON13] DAMER N., OPEL A., NOUAK A.: Performance anchored score normalization for multi-biometric fusion. In *9th International Symposium on Advances in Visual Computing, ISVC 2013, Rethymnon, Crete, Greece, July 29-31, 2013. Proceedings, Part II* (2013), Bebis G., Boyle R., Parvin B., Koracin D., Li B., Porikli F., Zordan V. B., Klosowski J. T., Coquillart S., Luo X., Chen M., Gotz D., (Eds.), vol. 8034 of *Lecture Notes in Computer Science*, Springer, pp. 68–75. 21, 41, 124

[DON14a] DAMER N., OPEL A., NOUAK A.: Biometric source weighting in multi-biometric fusion: Towards a generalized and robust solution. In *22nd European Signal Processing Conference, EUSIPCO 2014, Lisbon, Portugal, September 1-5, 2014* (2014), IEEE, pp. 1382–1386. 37, 56, 60, 89, 124

[DON14b] DAMER N., OPEL A., NOUAK A.: CMC curve properties and biometric source weighting in multi-biometric score-level fusion. In *17th International Conference on Information Fusion, FUSION 2014, Salamanca, Spain, July 7-10, 2014* (2014), IEEE, pp. 1–6. 37, 89, 124

[DOS13] DAMER N., OPEL A., SHAHVERDYAN A.: An overview on multi-biometric score-level fusion - verification and identification. In *ICPRAM 2013 - Proceedings of the 2nd International Conference on Pattern Recognition Applications and Methods, Barcelona, Spain, 15-18 February, 2013.* (2013), Marsico M. D., Fred A. L. N., (Eds.), SciTePress, pp. 647–653. 11, 21, 124

[DR10] DING Y., ROSS A.: When data goes missing: methods for missing score imputation in biometric fusion. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (2010), vol. 7667, p. 28. 23

[DR12] DING Y., ROSS A.: A comparison of imputation methods for handling missing scores in biometric fusion. *Pattern Recognition 45*, 3 (2012), 919 – 933. 23

[DRBK17] DAMER N., RHAIBANI C. I., BRAUN A., KUIJPER A.: Trust the biometrie mainstream: Multi-biometric fusion and score coherence. In *25th European Signal Processing Conference, EUSIPCO 2017, Kos, Greece, August 28 - September 2, 2017* (2017), IEEE, pp. 2191–2195. 55, 123

[DSN14] DAMER N., SAMARTZIDIS T., NOUAK A.: Personalized face reference from video: Key-face selection and feature-level fusion. In *Face and Facial Expression Recognition from Real World Videos - International Workshop, FFER@ICPR 2014, Stockholm, Sweden, August 24, 2014, Revised Selected Papers* (2014), Ji Q., Moeslund T. B., Hua G., Nasrollahi K., (Eds.), vol. 8912 of *Lecture Notes in Computer Science*, Springer, pp. 85–98. 85, 124

[DTBK17a] DAMER N., TERHÖRST P., BRAUN A., KUIJPER A.: Efficient, accurate, and rotation-invariant iris code. *IEEE Signal Processing Letters 24*, 8 (Aug 2017), 1233–1237. 71, 72, 73, 77, 123

[DTBK17b] DAMER N., TERHÖRST P., BRAUN A., KUIJPER A.: General borda count for multi-biometric retrieval. In *IEEE International Joint Conference on Biometrics, IJCB 2017, Denver, Colorado, USA, October 1 - 4, 2017* (2017), IEEE. 71, 123

[DTBK17c] DAMER N., TERHÖRST P., BRAUN A., KUIJPER A.: Indexing of single and multi-instance iris data based on lsh-forest and rotation invariant representation. In *Computer Analysis of Images and Patterns - 17th International Conference, CAIP 2017, Ystad, Sweden, August 22-24, 2017, Proceedings, Part II* (2017), Felsberg M., Heyden A., Krüger N., (Eds.), vol. 10425 of *Lecture Notes in Computer Science*, Springer, pp. 190–201. 71, 72, 73, 74, 123

[DWH*18] DAMER N., WAINAKH Y., HENNIGER O., CROLL C., BERTHE B., BRAUN A., KUIJPER A.: Deep learning-based face recognition and the robustness to perspective distortion. In *24th International Conference on Pattern Recognition, ICPR 2018, Beijing, China, August 2018* (2018), IEEE (under review). 126

[e-A15] E-AADHAAR - UNIQUE IDENTIFICATION AUTHORITY OF INDIA:. https://eaadhaar.uidai.gov.in/, 2015. 1, 71

[ES05] EKENEL H. K., STIEFELHAGEN R.: Local appearance based face recognition using discrete cosine transform. In *13th European Signal Processing Conference (EUSIPCO 2005* (2005). 88, 106

[Eub99] EUBANK R.: *Nonparametric Regression and Spline Smoothing, Second Edition*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1999. 31

[Far03] FARNEBÄCK G.: Two-frame motion estimation based on polynomial expansion. In *Image Analysis, 13th Scandinavian Conference, SCIA 2003, Halmstad, Sweden, June 29 - July 2, 2003, Proceedings* (2003), Bigün J., Gustavsson T., (Eds.), vol. 2749 of *Lecture Notes in Computer Science*, Springer, pp. 363–370. 101

[FKP08] FATUKASI O., KITTLER J., POH N.: Estimation of missing values in multimodal biometric fusion. In *Biometrics: Theory, Applications and Systems, 2008. BTAS 2008. 2nd IEEE International Conference on* (2008), pp. 1–6. 4, 23

[FR11] FRATRIC I., RIBARIC S.: Local binary lda for face recognition. In *Proceedings of the COST 2101 European conference on Biometrics and ID management* (Berlin, Heidelberg, 2011), BioID'11, Springer-Verlag, pp. 144–155. 86, 88, 106, 107, 108

[GAR10] GADDE R. B., ADJEROH D., ROSS A.: Indexing iris images using the burrows-wheeler transform. In *2010 IEEE International Workshop on Information Forensics and Security* (Dec 2010), pp. 1–6. 72

[GBP04] GYAOUROVA A., BEBIS G., PAVLIDIS I. T.: Fusion of infrared and visible images for face recognition. In *Computer Vision - ECCV 2004, 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part IV* (2004), Pajdla T., Matas J., (Eds.), vol. 3024 of *Lecture Notes in Computer Science*, Springer, pp. 456–468. 107

[GMAD05] GARCIA-SALICETTI S., MELLAKH M. A., ALLANO L., DORIZZI B.: Multimodal biometric score fusion: The mean rule vs. support vector classifiers. In *13th European Signal Processing Conference, EUSIPCO 2005, Antalya, Turkey, September 4-8, 2005* (2005), IEEE, pp. 1–4. 17

[GR09] GYAOUROVA A., ROSS A.: A coding scheme for indexing multimodal biometric databases. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops* (June 2009), pp. 93–98. 6, 72, 73, 74, 77

[GR12] GYAOUROVA A., ROSS A.: Index codes for multibiometric pattern retrieval. *IEEE Transactions on Information Forensics and Security 7*, 2 (April 2012), 518–529. 6, 72, 73, 74, 77

[GV00] GUTSCHOVEN B., VERLINDE P.: Multi-modal identity verification using support vector machines (svm). In *Information Fusion, 2000. FUSION 2000. Proceedings of the Third International Conference on* (July 2000), vol. 2, pp. THB3/3–THB3/8 vol.2. 56, 89

[GWL*13] GUAN G., WANG Z., LU S., DENG J. D., FENG D. D.: Keypoint-based keyframe selection. *IEEE Trans. Circuits Syst. Video Techn. 23*, 4 (2013), 729–734. 87

[Ham74] HAMPEL F. R.: The influence curve and its role in robust estimation. *Journal of the American Statistical Association 69*, 346 (1974), 383–393. 22

[HDB17] HENNIGER O., DAMER N., BRAUN A.: Opportunities for biometric technologies in smart environments. In *Ambient Intelligence - 13th European Conference, AmI 2017, Malaga, Spain, April 26-28, 2017, Proceedings* (2017), Braun A., Wichert R., Maña A., (Eds.), vol. 10217 of *Lecture Notes in Computer Science*, pp. 175–182. 124

[HDZ08] HAO F., DAUGMAN J., ZIELINSKI P.: A fast search algorithm for a large fuzzy database. *IEEE Trans. Information Forensics and Security 3*, 2 (2008), 203–212. 72

[HHS94] HO T. K., HULL J. J., SRIHARI S. N.: Decision combination in multiple classifier systems. *IEEE Transactions on Pattern Analysis and Machine Intelligence 16*, 1 (Jan 1994), 66–75. 74

[HMBLM08] HUANG G. B., MATTAR M., BERG T., LEARNED-MILLER E.: Labeled Faces in the Wild: A Database forStudying Face Recognition in Unconstrained Environments. In *Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition* (Marseille, France, 2008), Erik Learned-Miller and Andras Ferencz and Frédéric Jurie. 106

[HMM07] HUI H. P., MENG H. M., MAK M.: Adaptive weight estimation in multi-biometric verification using fuzzy logic decision fusion. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2007, Honolulu, Hawaii, USA, April 15-20, 2007* (2007), IEEE, pp. 501–504. 56

[HRRS86] HAMPEL F. R., RONCHETTI E. M., ROUSSEEUW P. J., STAHEL W. A.: *Robust Statistics: The Approach Based on Influence Functions*, 1st edition ed. Wiley Series in Probability and Statistics. Wiley, Jan. 1986. 23, 25

[HS12] HARIRI M., SHOKOUHI S. B.: Robustness of multi biometric authentication systems against spoofing. *Computer and Information Science* (2012), 77–86. 17

[HST07] HAO Y., SUN Z., TAN T.: Comparative studies on multispectral palm image fusion for biometrics. In *Computer Vision - ACCV 2007, 8th Asian Conference on Computer Vision, Tokyo, Japan, November 18-22, 2007, Proceedings, Part II* (2007), Yagi Y., Kang S. B., Kweon I., Zha H., (Eds.), vol. 4844 of *Lecture Notes in Computer Science*, Springer, pp. 12–21. 107

[Hue09] HUESKE E.: *Firearms and Fingerprints*. Facts on File/Infobase Publishing, New York, 2009. 11

[ICA*14] IRUJO J. A., CUADROS M., AGINAKO N., RAFFAELLI M., KÄHM O., DAMER N., NETO J. P.: Multimedia analysis of video sources. In *SIGMAP 2014 - Proceedings of the 11th International Conference on Signal Processing and Multimedia Applications, Vienna, Austria, 28-30 August, 2014* (2014), Obaidat M. S., Holzinger A., Cabello E., (Eds.), SciTePress, pp. 346–352. 125

[IKWY10] ICHINO M., KOMATSU N., WANG J.-G., YOU W. Y.: Speaker gender recognition using score level fusion by adaboost. In *11th International Conference on Control, Automation, Robotics and Vision, ICARCV 2010, Singapore, 7-10 December 2010, Proceedings* (2010), IEEE, pp. 648–

653. 17

[Int12] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: *ISO/IEC 2382-37:2012: Information technology Vocabulary Part 37: Biometrics*. Standard, ISO/IEC TC JTC1 SC37 Biometrics, Dec. 2012. 1, 11, 18

[Int17] INTERNATIONAL ORGANIZATION FOR STANDARDIZATION: "ISO/IEC DIS 30107-3:2017 information technology - biometric presentation attack detection - part 3: Testing and reporting. international organization for standardization", Sept. 2017. 102

[IR10] ISLAM M. R., RAHMAN M. F.: Article:likelihood ratio based score fusion for audio-visual speaker identification in challenging environment. *International Journal of Computer Applications 6*, 7 (September 2010), 6–11. Published By Foundation of Computer Science. 17

[JBP99] JAIN A., BOLLE R., PANKANTI S. (Eds.): *Biometrics : Personal Identification in Networked Society*. The Kluwer International Series in Engineering and Computer Science. Kluwer, 1999. 1, 11, 12

[JNR05] JAIN A., NANDAKUMAR K., ROSS A.: Score normalization in multimodal biometric systems. *Pattern Recognition 38*, 12 (2005), 2270 – 2285. 4, 5, 16, 21, 22, 23, 24, 26, 38, 57, 101

[JPDG08] JAYARAMAN U., PRAKASH S., DEVDATT, GUPTA P.: An indexing technique for biometric database. In *2008 International Conference on Wavelet Analysis and Pattern Recognition* (Aug 2008), vol. 2, pp. 758–763. 72

[JRN11] JAIN A., ROSS A., NANDAKUMAR K.: *Introduction to Biometrics*. Springer US, 2011. 86

[JRP04] JAIN A. K., ROSS A., PRABHAKAR S.: An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Techn. 14*, 1 (2004), 4–20. 13, 15

[JSH04] JIN A. T. B., SAMAD S. A., HUSSAIN A.: Nearest neighbourhood classifiers in a bimodal biometric verification system fusion decision scheme. *Journal of Research and Practice in Information Technology 36*, 1 (2004), 47–62. 17

[KD12a] KÄHM O., DAMER N.: 2d face liveness detection: An overview. In *2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group, Darmstadt, Germany, September 6-7, 2012* (2012), Brömme A., Busch C., (Eds.), vol. 197 of *LNI*, IEEE/GI, pp. 1–12. 125

[KD12b] KOSE N., DUGELAY J. L.: Classification of captured and recaptured images to detect photograph spoofing. In *2012 International Conference on Informatics, Electronics Vision, ICIEV 2012,* (May 2012), pp. 1027–1032. 105

[KFFB07] KOLLREIDER K., FRONTHALER H., FARAJ M. I., BIGUN J.: Real-time face detection and motion analysis with application in "liveness" assessment. *IEEE Transactions on Information Forensics and Security 2*, 3 (Sept 2007), 548–558. 99

[KG00] KLOSTERMAN A. J., GANGER G. R.: Secure continuous biometric-enhanced authentication. *Research Showcase @ Carnegie Mellon University* (2000). 88

[kJuJhY06] KEUN JEE H., UK JUNG S., HEE YOO J.: Liveness detection for embedded face recognition system. *International Journal of Biomedical Sciences* (2006), 235–238. 99

[KNR09] K. NANDAKUMAR A. K. J., ROSS A.: Fusion in multibiometric identification systems: What about the missing data? In *Proceedings of the 3rd International Conference on Biometrics, Alghero, Italy* (June 2009). 14, 15, 17

[Kre00] KREYSZIG E.: *Advanced Engineering Mathematics: Maple Computer Guide*, 8th ed. John Wiley & Sons, Inc., New York, NY, USA, 2000. 22

[KTT10] KIM Y., TOH K., TEOH A. B. J.: An online learning algorithm for biometric scores fusion. In *Fourth IEEE International Conference on Biometrics: Theory Applications and Systems, BTAS 2010, Washington, DC, USA, 27-29 September, 2010* (2010), IEEE, pp. 1–6. 14, 15

[LdC10] LORENA A. C., DE CARVALHO A. C. P. L. F.: Building binary-tree-based multiclass classifiers using separability measures. *Neurocomput. 73*, 16-18 (Oct. 2010), 2837–2845. 39

[LHL*13] LI H., HUA G., LIN Z., BRANDT J., YANG J.: Probabilistic elastic matching for pose variant face verification. *2013 IEEE Conference on Computer Vision and Pattern Recognition 0* (2013), 3499–3506. 106, 112

[Lit92] LITTLE R. J. A.: Regression With Missing X's: A Review. *Journal of the American Statistical Association 87*, 420 (1992), 1227–1237. 4, 23

[LJ11] LI S. Z., JAIN A. K.: *Handbook of Face Recognition*, 2nd ed. Springer Publishing Company, Incorporated, 2011. 20, 86

[LL11] L. LATHA S. T.: On improving the performance of multimodal biometric authentication through ant colony optimization. *WSEAS Transactions on Information Science and Applications* (2011). 17

[LLK*13] LEYS C., LEY C., KLEIN O., BERNARD P., LICATA L.: Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology 49*, 4 (2013), 764 – 766. 22

[Low04] LOWE D.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision 60*, 2 (2004), 91–110. 88, 106

[LPV03a] LU J., PLATANIOTIS K. N., VENETSANOPOULOS A. N.: Face recognition using lda-based algorithms. *IEEE Trans. Neural Networks 14*, 1 (2003), 195–200. 88, 106, 107

[LPV03b] LU J., PLATANIOTIS K. N., VENETSANOPOULOS A. N.: Regularized discriminant analysis for the small sample size problem in face recognition. *Pattern Recogn. Lett. 24*, 16 (Dec. 2003), 3079–3087. 108

[MB13] MONDAL S., BOURS P.: Continuous authentication using mouse dynamics. In *2013 BIOSIG - Proceedings of the 12th International Conference of Biometrics Special Interest Group, Darmstadt, Germany, September 4-6, 2013* (2013), Brömme A., Busch C., (Eds.), vol. 212 of *LNI*, GI, pp. 123–134. 85

[MB15] MONDAL S., BOURS P.: Continuous authentication in a real world settings. In *Eighth International Conference on Advances in Pattern Recognition, ICAPR 2015, Kolkata, India, January 4-7, 2015* (2015), IEEE, pp. 1–6. 89

[MD15] MAUL F., DAMER N.: Fuzzy logic and multi-biometric fusion - an overview. In *ICPRAM 2015 - Proceedings of the International Conference on Pattern Recognition Applications and Methods, Volume 1, Lisbon, Portugal, 10-12 January, 2015.* (2015), Marsico M. D., Figueiredo M. A. T., Fred A. L. N., (Eds.), SciTePress, pp. 218–222. 124

[MHP11] MÄÄTTÄ J., HADID A., PIETIKÄINEN M.: Face spoofing detection from single images using micro-texture analysis. In *2011 IEEE International Joint Conference on Biometrics, IJCB 2011, Washington, DC, USA, October 11-13, 2011* (2011), IEEE, pp. 1–7. 99, 105

[Mit12] MITCHELL H. B.: *Data Fusion Concepts and Ideas; 2nd ed.* Springer, Berlin, 2012. 13

[MNL14] MARCEL S., NIXON M. S., LI S. Z.: *Handbook of Biometric Anti-Spoofing: Trusted Biometrics Under Spoofing Attacks*. Springer Publishing Company, Incorporated, 2014. 86

[MP09] MOIN M. S., PARVIZ M.: Exploring auc boosting approach in multimodal biometrics score level fusion. In *Proceedings of the 2009 Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing* (Washington, DC, USA, 2009), IIH-MSP '09, IEEE Computer Society, pp. 616–619. 17

[MR08] MUKHERJEE R., ROSS A.: Indexing iris images. In *19th International Conference on Pattern Recognition (ICPR 2008), December 8-11, 2008, Tampa, Florida, USA* (2008), IEEE Computer Society, pp. 1–4. 72

[MS05] MIKOLAJCZYK K., SCHMID C.: A performance evaluation of local descriptors. *IEEE Trans. Pattern Anal. Mach. Intell. 27*, 10 (Oct. 2005), 1615–1630. 58

[MS08] MIMAROGLU S., SIMOVICI D. A.: Approximate computation of object distances by locality-sensitive hashing. In *Proceedings of The 2008 International Conference on Data Mining, DMIN 2008, July 14-17, 2008, Las Vegas, USA, 2 Volumes* (2008), Stahlbock R., Crone S. F., Lessmann S., (Eds.), CSREA Press, pp. 714–718. 76

[MSMG09] MEHROTRA H., SRINIVAS B. G., MAJHI B., GUPTA P.: Indexing iris biometric database using energy histogram of DCT subbands. In *Contemporary Computing - Second International Conference, IC3 2009, Noida, India, August 17-19, 2009. Proceedings* (2009), Ranka S., Aluru S., Buyya R., Chung Y., Dua S., Grama A., Gupta S. K. S., Kumar R., Phoha V. V., (Eds.), vol. 40 of *Communications in Computer and Information Science*, Springer, pp. 194–204. 72

[MSV99] MORENO B., SANCHEZ A., VELEZ J.: On the use of outer ear images for personal identification in security applications. In *IEEE 33rd Annual 1999 International Carnahan Conference on Security Technology, 1999* (1999), pp. 469 –476. 15

[MVSM12] MEHROTRA H., VATSA M., SINGH R., MAJHI B.: Biometric match score fusion using RVM: A case study in multi-unit iris recognition. In *2012 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Providence, RI, USA, June 16-21, 2012* (2012), IEEE Computer Society, pp. 65–70. 17

[NCDJ08] NANDAKUMAR K., CHEN Y., DASS S. C., JAIN A.: Likelihood ratio-based biometric score fusion. *IEEE Trans. Pattern Anal. Mach. Intell. 30*, 2 (Feb. 2008), 342–347. 17, 56, 89

[NCJD06] NANDAKUMAR K., CHEN Y., JAIN A. K., DASS S. C.: Quality-based score level fusion in multibiometric systems. In *Proceedings of the 18th International Conference on Pattern Recognition - Volume 04* (Washington, DC, USA, 2006), ICPR '06, IEEE Computer Society, pp. 473–476. 5, 17, 56

[NIS] NIST: *National Institute of Standards and Technology: NIST Biometric Scores Set*. 41, 48, 55, 58, 59, 61

[NJR09] NANDAKUMAR K., JAIN A. K., ROSS A.: Fusion in multibiometric identification systems: What about the missing data? In *Proceedings of the Third International Conference on Advances in Biometrics* (Berlin, Heidelberg, 2009), ICB '09, Springer-Verlag, pp. 743–752. 23

[NPJ10] NIINUMA K., PARK U., JAIN A. K.: Soft biometric traits for continuous user authentication. *IEEE Trans. Information Forensics and Security 5*, 4 (2010), 771–780. 88

[NS09] NISHA SRINIVAS KALYAN VEERAMACHANENI L. O.: Fusing correlated data from multiple classifiers for improved biometric verification. In *12th International Conference on Information Fusion* (2009). 16, 24

[OFA*10] ORTEGA-GARCIA J., FIÉRREZ J., ALONSO-FERNANDEZ F., GALBALLY J., FREIRE M. R., GONZALEZ-RODRIGUEZ J., GARCÍA-MATEO C., ALBA-CASTRO J. L., GONZÁLEZ-

AGULLA E., MURAS E. O., GARCIA-SALICETTI S., ALLANO L., LY V., DORIZZI B., KITTLER J., BOURLAI T., POH N., DERAVI F., NG M. W. R., FAIRHURST M. C., HENNEBERT J., HUMM A., TISTARELLI M., BRODO L., RICHIARDI J., DRYGAJLO A., GANSTER H., SUKNO F., PAVANI S., FRANGI A. F., AKARUN L., SAVRAN A.: The multiscenario multienvironment biosecure multimodal database (BMDB). *IEEE Trans. Pattern Anal. Mach. Intell. 32*, 6 (2010), 1097–1111. 56, 65

[OPH96] OJALA T., PIETIKÄINEN M., HARWOOD D.: A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition 29*, 1 (Jan. 1996), 51–59. 88, 106

[PA07] PROENCA H., ALEXANDRE L. A.: Toward noncooperative iris recognition: A classification approach using multiple signatures. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 4 (Apr. 2007), 607–612. 71

[PB03] POH N., BENGIO S.: Non-linear variance reduction techniques in biometric authentication. In *In Workshop on Multimodal User Authentication (MMUA) 2003* (2003), pp. 123–130. 38

[PB04] POH N., BENGIO S.: *A Study of the Effects of Score Normalisation Prior to Fusion in Biometric Authentication Tasks*. Idiap-RR Idiap-RR-69-2004, IDIAP, 0 2004. 5, 38, 39

[PB06] POH N., BENGIO S.: Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication. *Pattern Recognition 39*, 2 (2006), 223 – 233. 21, 26, 27, 28, 41

[PBK*09] POH N., BOURLAI T., KITTLER J., ALLANO L., ALONSO-FERNANDEZ F., AMBEKAR O., BAKER J., DORIZZI B., FATUKASI O., FIERREZ J., GANSTER H., ORTEGA-GARCIA J., MAURER D., SALAH A., SCHEIDAT T., VIELHAUER C.: Benchmarking quality-dependent and cost-sensitive score-level multimodal biometric fusion algorithms. *Information Forensics and Security, IEEE Transactions on 4*, 4 (Dec 2009), 849–866. 38

[PBK10] POH N., BOURLAI T., KITTLER J.: A multimodal biometric test bed for quality-dependent, cost-sensitive and client-specific score-level fusion algorithms. *Pattern Recogn. 43*, 3 (Mar. 2010), 1094–1105. 23, 31, 41

[PCK*12] POH N., CHAN C.-H., KITTLER J., FIERREZ J., , GALBALLY J.: *D3.3: Description of Metrics For the Evaluation of Biometric Performance*. Deliverable, FP7 project: BEAT Biometrics Evaluation and Testing, August 2012. 19

[PDC09] PINTO N., DICARLO J. J., COX D. D.: How far can you get with a modern face recognition test set using only simple features? In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009), 20-25 June 2009, Miami, Florida, USA* (2009), IEEE Computer Society, pp. 2591–2598. 107

[PJ01] PRABHAKAR S., JAIN A.: Decision-level fusion in fingerprint verification. In *Multiple Classifier Systems*, Kittler J., Roli F., (Eds.), vol. 2096 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2001, pp. 88–98. 107

[PK08] POH N., KITTLER J.: A family of methods for quality-based multimodal biometric fusion using generative classifiers. In *10th International Conference on Control, Automation, Robotics and Vision, ICARCV 2008, Hanoi, Vietnam, 17-20 December 2008, Proceedings* (2008), IEEE, pp. 1162–1167. 5, 17, 56

[PMB07] POH N., MARTIN A., BENGIO S.: Performance generalization in biometric authentication using joint user-specific and sample bootstraps. *IEEE Trans. Pattern Anal. Mach. Intell. 29*, 3 (2007), 492–498. 45

[PMK09] POH N., MERATI A., KITTLER J.: Making better biometric decisions with quality and cohort information: A case study in fingerprint verification. In *17th European Signal Processing Conference, EUSIPCO 2009, Glasgow, Scotland, UK, August 24-28, 2009* (2009), IEEE, pp. 70–74. 5, 17, 56

[PWM*10] POH N., WINDRIDGE D., MOTTL V., TATARCHUK A., ELISEYEV A.: Addressing missing values in kernel-based multimodal biometric fusion using neutral point substitution. *IEEE Trans. Information Forensics and Security 5*, 3 (2010), 461–469. 4, 17, 23

[RBBB14] RATHGEB C., BREITINGER F., BUSCH C., BAIER H.: On application of bloom filters to iris biometrics. *IET Biometrics 3*, 4 (2014), 207–218. 72

[RBBB15] RATHGEB C., BREITINGER F., BAIER H., BUSCH C.: Towards bloom filter-based indexing of iris biometric data. In *2015 International Conference on Biometrics (ICB)* (May 2015), pp. 422–429. 72

[RCLM08] RODRÍGUEZ L. P., CRESPO A. G., LARA M. J. P., MEZCUA B. R.: Study of different fusion techniques for multimodal biometric authentication. In *Proceedings of the 2008 IEEE International Conference on Wireless & Mobile Computing, Networking & Communication* (Washington, DC, USA, 2008), WIMOB '08, IEEE Computer Society, pp. 666–671. 17

[RDRK11] RAGHAVENDRA R., DORIZZI B., RAO A., KUMAR G. H.: Designing efficient fusion schemes for multimodal biometric systems using face and palmprint. *Pattern Recognition 44*, 5 (2011), 1076 – 1088. 17, 107

[Rho56] RHODES H. T. F.: *Alphonse Bertillon, father of scientific detection*. Abelard-Schuman, 1956. 11

[RJ03] ROSS A., JAIN A. K.: Information fusion in biometrics. *Pattern Recognition Letters 24*, 13 (2003), 2115–2125. 5, 23, 37, 89

[RKBT06] RATTANI A., KISKU D. R., BICEGO M., TISTARELLI M.: Robust Feature-Level Multibiometric Classification. *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the* (2006), 1–6. 14

[RNJ06] ROSS A. A., NANDAKUMAR K., JAIN A. K.: *Handbook of Multibiometrics (International Series on Biometrics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. 17

[Rom11] ROMERO J. J.: Fast start for world's biggest biometrics id project in india, a few million people have been registered for a biometric database so far - only a billion left to go. *IEEE Spectrum* (May 2011). 1

[RU10] RATHGEB C., UHL A.: Iris-biometric hash generation for biometric database indexing. In *20th International Conference on Pattern Recognition, ICPR 2010, Istanbul, Turkey, 23-26 August 2010* (2010), IEEE Computer Society, pp. 2848–2851. 72

[Rub87] RUBIN D. B.: *Multiple Imputation for Nonresponse in Surveys*. Wiley, 1987. 23

[SB14] SOUSEDIK C., BUSCH C.: Presentation attack detection methods for fingerprint recognition systems: a survey. *IET Biometrics 3*, 4 (2014), 219–233. 86

[SBCI10] SOLAMI E. A., BOYD C., CLARK A. J., ISLAM A. K.: Continuous biometric authentication: Can it be more practical? In *12th IEEE International Conference on High Performance Computing and Communications, HPCC 2010, 1-3 September 2010, Melbourne, Australia* (2010), IEEE, pp. 647–652. 88

[SC08] SLANEY M., CASEY M.: Locality-Sensitive Hashing for Finding Nearest Neighbors. *IEEE Signal Processing Magazine 25*, 2 (Mar. 2008), 128–131. 74

[SCWT14] SUN Y., CHEN Y., WANG X., TANG X.: Deep learning face representation by joint identification-verification. In *Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13 2014, Montreal, Quebec, Canada* (2014), Ghahramani Z., Welling M., Cortes C., Lawrence N. D., Weinberger K. Q., (Eds.), pp. 1988–1996. 106

[SED16] SIEGMUND D., EBERT T., DAMER N.: Combining low-level features of offline questionnaires for handwriting identification. In *Image Analysis and Recognition - 13th International Conference, ICIAR 2016, in Memory of Mohamed Kamel, Póvoa de Varzim, Portugal, July 13-15, 2016, Proceedings* (2016), Campilho A., Karray F., (Eds.), vol. 9730 of *Lecture Notes in Computer Science*, Springer, pp. 46–54. 124

[SG07] SINGH Y. N., GUPTA P.: Quantitative evaluation of normalization techniques of matching scores in multimodal biometric systems. In *Proceedings of the 2007 international conference on Advances in Biometrics* (Berlin, Heidelberg, 2007), ICB'07, Springer-Verlag, pp. 574–583. 22, 23

[SKP15] SCHROFF F., KALENICHENKO D., PHILBIN J.: Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2015), pp. 815–823. 65

[SS99] SCHAPIRE R. E., SINGER Y.: Improved boosting algorithms using confidence-rated predictions. *Machine Learning 37*, 3 (1999), 297–336. 86, 101

[SSD*14] SIEGMUND D., SAMARTZIDIS T., DAMER N., NOUAK A., BUSCH C.: Virtual fitting pipeline: Body dimension recognition, cloth modeling, and on-body simulation. In *VRIPHYS 14: 11th Workshop on Virtual Reality Interactions and Physical Simulations, Bremen, Germany, 2014. Proceedings* (2014), Bender J., Duriez C., Jaillet F., Zachmann G., (Eds.), Eurographics Association, pp. 99–107. 125

[SSG*18] SAMARTZIDIS T., SIEGMUND D., GOEDDE M., DAMER N., BRAUN A., KUIJPER A.: The dark side of the face: exploring the ultraviolet spectrum for face biometrics. In *International Conference on Biometrics, ICB 2018, 20-23 February, 2018, Gold Coast, Queensland, Australia* (2018), IEEE (to appear). 124

[ST09] SUN Z., TAN T.: Ordinal measures for iris recognition. *IEEE Trans. Pattern Anal. Mach. Intell. 31*, 12 (2009), 2211–2226. 72, 73, 77

[SUM*05] SNELICK R., ULUDAG U., MINK A., INDOVINA M., JAIN A. K.: Large-scale evaluation of multimodal biometric authentication using state-of-the-art systems. *IEEE Trans. Pattern Anal. Mach. Intell. 27*, 3 (2005), 450–455. 5, 22, 23, 38

[SVN07] SINGH R., VATSA M., NOORE A.: Intelligent biometric information fusion using support vector machine. In *Soft Computing in Image Processing*, Nachtegael M., Van der Weken D., Kerre E., Philips W., (Eds.), vol. 210 of *Studies in Fuzziness and Soft Computing*. Springer Berlin Heidelberg, 2007, pp. 325–349. 17, 56, 89

[SWT14] SUN Y., WANG X., TANG X.: Deep learning face representation from predicting 10, 000 classes. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014* (2014), IEEE Computer Society, pp. 1891–1898. 106

[SZL*11] SONG S., ZHAN Z., LONG Z., ZHANG J., YAO L.: Comparative study of svm methods combined with voxel selection for object category classification on fmri data. *PLoS ONE 6*, 2 (02 2011), e17191. 58

[TDB*18] TERHÖRST P., DAMER N., BOUTROS F., BRAUN A., KUIJPER A.: What can a single minutia tell about gender? In *6th IAPR/IEEE International Workshop on Biometrics and Forensics, Sassari, IT, June 7, 8 2018* (2018), IEEE (under review). 126

[TKL08] TOH K.-A., KIM J., LEE S.: Maximizing area under roc curve for biometric scores fusion. *Pattern Recogn. 41*, 11 (Nov. 2008), 3373–3392. 5, 38

[TLLJ10] TAN X., LI Y., LIU J., JIANG L.: Face liveness detection from a single image with sparse low rank bilinear discriminative model. In *Computer Vision - ECCV 2010 - 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010, Proceedings, Part VI* (2010), Daniilidis K., Maragos P., Paragios N., (Eds.), vol. 6316 of *Lecture Notes in Computer Science*, Springer, pp. 504–517. 86

[TLM08] TSAI D., LEE Y., MATSUYAMA E.: Information entropy measure for evaluation of image quality. *J. Digital Imaging 21*, 3 (2008), 338–347. 108

[TWL10] TONG Y., WHEELER F. W., LIU X.: Improving biometric identification through quality-based face and fingerprint biometric fusion. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2010, San Francisco, CA, USA, 13-18 June, 2010* (2010), IEEE Computer Society, pp. 53–60. 14

[TYRW14] TAIGMAN Y., YANG M., RANZATO M., WOLF L.: Deepface: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014* (2014), IEEE Computer Society, pp. 1701–1708. 106

[Vap95] VAPNIK V. N.: *The Nature of Statistical Learning Theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995. 58

[Ver13] *2013 Data Breach Investigations Report*. Technical report, Verizon RISK team, Dec. 2013. 1

[VIM*07] VAJARIA H., ISLAM T., MOHANTY P., SARKAR S., SANKAR R., KASTURI R.: Evaluation and analysis of a face and voice outdoor multi-biometric system. *Pattern Recogn. Lett. 28*, 12 (Sept. 2007), 1572–1580. 16, 24

[VJ01] VIOLA P. A., JONES M. J.: Rapid object detection using a boosted cascade of simple features. In *2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), with CD-ROM, 8-14 December 2001, Kauai, HI, USA* (2001), IEEE Computer Society, pp. 511–518. 101, 108

[VMC13] VAZQUEZ H. M., MARTÍNEZ-DÍAZ Y., CHAI Z.: Volume structured ordinal features with background similarity measure for video face recognition. In *International Conference on Biometrics, ICB 2013, 4-7 June, 2013, Madrid, Spain* (2013), Fiérrez J., Kumar A., Vatsa M., Veldhuis R. N. J., Ortega-Garcia J., (Eds.), IEEE, pp. 1–6. 107, 112

[VP09] VILLEGAS M., PAREDES R.: Score fusion by maximizing the area under the ROC curve. In *Pattern Recognition and Image Analysis, 4th Iberian Conference, IbPRIA 2009, Póvoa de Varzim, Portugal, June 10-12, 2009, Proceedings* (2009), Araújo H., Mendonça A. M., Pinho A. J., Torres M. I., (Eds.), vol. 5524 of *Lecture Notes in Computer Science*, Springer, pp. 473–480. 5, 38

[VT12] VIRIRI S., TAPAMO J.: Integrating iris and signature traits for personal authentication using user-specificweighting. *Sensors 12*, 4 (2012), 4324–4338. 38

[WGT*07] WATSON C., GARRIS M., TABASSI E., WILSON C., MCCABE R., JANET S., KO K., OF STANDARDS N. I., (U.S.) T.: *User's Guide to NIST Biometric Image Software (NBIS)*.

2007. 66

[WHJ15] WEN D., HAN H., JAIN A. K.: Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security 10*, 4 (April 2015), 746–761. 7, 86, 99, 103, 105

[WHM11] WOLF L., HASSNER T., MAOZ I.: Face recognition in unconstrained videos with matched background similarity. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (June 2011), pp. 529–534. 87, 106, 109, 111

[WHT10] WOLF L., HASSNER T., TAIGMAN Y.: Similarity scores based on background samples. In *Proceedings of the 9th Asian Conference on Computer Vision - Volume Part II* (Berlin, Heidelberg, 2010), ACCV'09, Springer-Verlag, pp. 88–97. 88, 106

[WL13] WOLF L., LEVY N.: The svm-minus similarity score for video face recognition. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013* (2013), IEEE Computer Society, pp. 3523–3530. 106

[WTJ03] WANG Y., TAN T., JAIN A.: Combining face and iris biometrics for identity verification. In *Audio- and Video-Based Biometric Person Authentication*, Kittler J., Nixon M., (Eds.), vol. 2688 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2003, pp. 805–813. 14, 107

[WTS08] WEI Z., TAN T., SUN Z.: Synthesis of large realistic iris databases using patch-based sampling. In *Pattern Recognition, 2008. ICPR 2008. 19th International Conference on* (Dec 2008), pp. 1–4. 77

[WWG*07] WU Q., WANG L., GENG X., LI M., HE X.: Dynamic biometrics fusion at feature level for video- based human recognition. In *Proceedings of Image and Vision Computing New Zealand 2007, pp. 152ï£¡157, Hamilton, New Zealand, December 2007* (2007). 56

[Yan06] YAN P.: *Ear biometrics in human identification*. PhD thesis, University of Notre Dame, Notre Dame, IN, USA, 2006. AAI3406950. 14

[YB05] YAN P., BOWYER K. W.: Multi-biometrics 2d and 3d ear recognition. In *5th International Conference on Audio- and Video-Based Biometric Person Authentication, AVBPA 2005, Hilton Rye Town, NY, USA, July 20-22, 2005, Proceedings* (2005), Kanade T., Jain A. K., Ratha N. K., (Eds.), vol. 3546 of *Lecture Notes in Computer Science*, Springer, pp. 503–512. 15

[YLHZ12] YANG Y., LIN K., HAN F., ZHANG Z.: Dynamic weighting for effective fusion of fingerprint and finger vein. In *Progress in Intelligent Computing and Applications Volume1, Number1,October 2012* (2012). 56

[YSKR08] YAP R. H., SIM T., KWANG G. X., RAMNATH R.: Physical access protection using continuous authentication. In *2008 IEEE Conference on Technologies for Homeland Security* (2008), IEEE, pp. 510–512. 88

[ZD15] ZHONG Y., DENG Y.: *Recent Advances in User Authentication Keystroke Dynamics Biometrics*. Science Gate Publishing, 2015, ch. A Survey on Keystroke Dynamics Biometrics: Approaches, Advances, and Evaluations, pp. 1–22. 7, 85, 89

[ZHHD16] ZHANG N., HAN J., HU J., DENG W.: Locally rejected metric learning based false positives filtering for face detection. In *11th Chinese Conference on Biometric Recognition, CCBR 2016, Chengdu, China, October 14-16, 2016, Proceedings* (2016), You Z., Zhou J., Wang Y., Sun Z., Shan S., Zheng W., Feng J., Zhao Q., (Eds.), vol. 9967 of *Lecture Notes in Computer Science*, pp. 13–21. 108

[ZYL*12] ZHANG Z., YAN J., LIU S., LEI Z., YI D., LI S. Z.: A face antispoofing database with diverse attacks. In *2012 5th IAPR International Conference on Biometrics (ICB)* (March 2012), pp. 26–31. 86, 99, 105